

GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on:

March 18, 2003, 15:29:05 ; Search time 40 Seconds

(without alignments)  
2864.888 Million cell updates/sec

Title: US-09-781-693a-2

Perfect score: 4527

Sequence: 1 MSRGSSYPHLLMDVKKRSLG.....RLGDRSGSCQENENDEE 860

Scoring table:

BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched:

908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters:

908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database:

Listing first 120 summaries

1. /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2. /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3. /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4. /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5. /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6. /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7. /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8. /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9. /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10. /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11. /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12. /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13. /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14. /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15. /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16. /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17. /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18. /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19. /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20. /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21. /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22. /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23. /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	4514	99.7	860	22 AAB69556 Human Repro-EN-1.0
2	4498	99.4	880	22 AAB69505 Human cell cycle a
3	4455.5	98.6	837	22 AAB69557 Human IBI protein.
4	2390	52.8	482	23 AAB05707 Human signal trans.
5	1459	32.2	301	22 AAU19599 Human diagnostic a
6	847	18.7	163	22 AAB92537 Human protein sequ
7	688	15.2	129	21 AAB27984 Human secreted pro
8	664	14.7	127	22 AAB27926 Human secreted pro
9	569	12.6	110	22 AAB29956 Peptide #2607 enco
10	569	12.6	110	22 AAB20555 Protein #2554 enco

11	569	12.6	110	22 AAM5960 Human brain expres
12	569	12.6	110	22 AAM68328 Human bone marrow
13	569	12.6	110	22 AAM16144 Peptide #2578 enco
14	569	12.6	110	22 AAM03874 Peptide #2556 enco
15	569	12.6	110	22 AAB37880 Human peptide enco
16	508.5	11.2	628	23 AAB61280 Drosophila melanog
17	508.5	11.2	628	23 AAB80462 Fruit fly adipose
18	499.5	11.0	677	23 AAB80463 Human adipose prot
19	489.5	10.6	677	23 AAB80464 Human adipose prot
20	481.5	10.4	536	23 AAB80470 Human adipose prot
21	470.5	10.2	580	23 AAB80471 Human adipose prot
22	460.5	10.2	580	23 AAB80472 Mouse adipose prot
23	441	9.7	616	23 AAB80473 Mouse adipose prot
24	437	9.7	86	22 AAB27840 Human adipose prot
25	437	9.7	86	22 AAB33011 Human adipose prot
26	437	9.7	86	22 AAB18484 Peptide #517 enco
27	437	9.7	86	22 AAM33809 Protein #483 enco
28	437	9.7	86	22 AAM6195 Human bone marrow
29	437	9.7	86	22 AAM26472 Peptide #509 enco
30	437	9.7	86	22 AAM01806 Peptide #488 enco
31	437	9.7	86	22 AAM14066 Human peptide enco
32	437	9.7	86	22 AAG18887 Arabidopsis thalia
33	436.5	9.6	471	21 AAG50441 Arabidopsis thalia
34	436.5	9.6	471	21 AAG50441 Arabidopsis thalia
35	429.5	9.5	816	16 AAB85870 Novel human diagno
36	404	8.9	633	22 AAB27224 Novel human diagno
37	399	8.8	385	23 AAB57841 Fruit fly adipose
38	373	8.2	748	22 AAB27841 Drosophila melanog
39	338	7.5	62	22 AAB33012 Peptide #518 enco
40	338	7.5	62	22 AAB18485 Protein #484 enco
41	338	7.5	62	22 AAM65810 Human brain expres
42	338	7.5	62	22 AAM14067 Human bone marrow
43	338	7.5	62	22 AAM26473 Peptide #501 enco
44	338	7.5	62	22 AAB35846 Peptide #489 enco
45	338	7.5	62	22 AAB93275 Human peptide enco
46	338	7.5	62	22 AAB94430 Human protein sequ
47	338	7.5	62	22 AAG18389 Arabidopsis thalia
48	330	7.3	576	23 AAG08447 Arabidopsis thalia
49	329	6.8	60	22 AAB69558 Human Repro-EN-1.0
50	306	6.5	60	22 AAB69557 Human Repro-EN-1.0
51	296	6.0	214	21 AAB57007 Human prostate can
52	273	6.0	365	21 AAG18388 Arabidopsis thalia
53	269.5	6.0	365	21 AAG50442 Arabidopsis thalia
54	269.5	6.0	365	21 AAB94430 Human secreted pro
55	264	5.8	129	21 AAB94430 Human protein sequ
56	245.5	5.4	355	22 AAG18389 Arabidopsis thalia
57	244.5	5.3	355	22 AAG50443 Arabidopsis thalia
58	238.5	5.3	258	21 AAG08448 Arabidopsis thalia
59	238.5	5.2	191	21 AAG08448 Arabidopsis thalia
60	234.5	5.2	191	21 AAG08448 Arabidopsis thalia
61	234.5	5.2	191	21 AAG08448 Arabidopsis thalia
62	234	5.2	191	21 AAG08448 Arabidopsis thalia
63	232.5	5.1	192	22 AAM95643 Human immune/thema
64	230	5.1	192	22 AAM95643 Human polypeptide
65	214	4.7	174	21 AAG08449 Arabidopsis thalia
66	213.5	4.7	174	21 AAG08449 Arabidopsis thalia
67	205.5	4.5	1301	22 AAB67528 Human Repro-EN-1.0
68	204.5	4.5	1219	22 AAB67528 Human Repro-EN-1.0
69	204.5	4.5	1219	22 AAB67528 Human Repro-EN-1.0
70	203	4.5	1219	22 AAB67528 Human Repro-EN-1.0
71	201	4.4	1219	22 AAB67528 Human Repro-EN-1.0
72	200	4.4	1219	22 AAB67528 Human Repro-EN-1.0
73	200	4.4	1219	22 AAB67528 Human Repro-EN-1.0
74	195.5	4.3	1162	22 AAB62516 Drosophila melanog
75	194.5	4.3	1162	22 AAB62516 Drosophila melanog
76	194.5	4.3	1162	22 AAB62516 Drosophila melanog
77	193	4.3	1162	22 AAB62516 Drosophila melanog
78	192.5	4.3	1162	22 AAB62516 Drosophila melanog
79	192.5	4.3	1162	22 AAB62516 Drosophila melanog
80	191.5	4.2	985	16 AAM0634 Novel human secret
81	191	4.2	985	16 AAM0634 Novel human secret
82	189	4.2	2951	22 AAB60786 IIV glycoprotein
83	189	4.2	2951	22 AAB60786 IIV glycoprotein
84	189	4.2	2951	22 AAB60786 IIV glycoprotein
85	189	4.2	2951	22 AAB60786 IIV glycoprotein
86	189	4.2	2951	22 AAB60786 IIV glycoprotein
87	189	4.2	2951	22 AAB60786 IIV glycoprotein
88	189	4.2	2951	22 AAB60786 IIV glycoprotein
89	189	4.2	2951	22 AAB60786 IIV glycoprotein
90	189	4.2	2951	22 AAB60786 IIV glycoprotein
91	189	4.2	2951	22 AAB60786 IIV glycoprotein
92	189	4.2	2951	22 AAB60786 IIV glycoprotein
93	189	4.2	2951	22 AAB60786 IIV glycoprotein
94	189	4.2	2951	22 AAB60786 IIV glycoprotein
95	189	4.2	2951	22 AAB60786 IIV glycoprotein
96	189	4.2	2951	22 AAB60786 IIV glycoprotein
97	189	4.2	2951	22 AAB60786 IIV glycoprotein
98	189	4.2	2951	22 AAB60786 IIV glycoprotein
99	189	4.2	2951	22 AAB60786 IIV glycoprotein
100	189	4.2	2951	22 AAB60786 IIV glycoprotein

Amnio acid sequenc  
Novel human diagn  
Novel human diagn  
Novel human diagn  
Human polypeptide  
Amnio acid sequenc  
Novel human diagn  
Novel human diagn  
Novel human diagn  
Streptococcus pau  
S. pneumoniae SpD  
Human ORFX OR2/760  
Drosophila melanog  
Novel human diagn  
Staphylococcus aur  
Human zci protein.  
Novel human diagn  
Human HPK/GSK-like  
Human kinase /PRIN  
Human kinase  
Novel human diagn  
Plasmodium falcipla  
Human Repro-FN-1.0  
Peptide #2316 enco  
Peptide #2403 enco  
Protein #2313 enco  
Human bone mario  
Human brain mario  
Peptide #2327 enco  
Human peptide enco  
Lactococcus lactis  
Drosophila melanog  
Novel human diagn  
Staphylococcus aur  
Staphylococcus aur  
Mouse SPRE20-related  
Human ORFX ORF289  
Drosophila melanog  
Human dentin matir

RESULT 1  
AAB69556 standard; Protein; 860 AA.

27-APR-2001 (first entry)  
Human Repro-EN-1.0 protein.  
Human; Repro-EN-1.0; IB1; cytosol; breast cancer; uterine cancer  
prostate cancer; epitope.

AL.

Just as you

XX (DIAG-) DIAGNOSTIC PROD CORP.  
PA  
XX  
XX El Shami AS, Menon SN, French CK  
XX  
XX WPI: 2001-182795/18.  
DR N-PSDB; AAF58654.

New autoantigens Repro-EW-1.0 and IB 1 polypeptides and polynucleotides are useful for diagnosing endometriosis or as a marker for pathologic conditions such as breast, uterine or prostate cancer.

The present sequence is given in a specification relating to recombinant polynucleotides comprising nucleotide sequences encoding a polypeptide epitope of at least 5 amino acids of Repro-EN-1.0 or of IB 1 having a fully defined sequence of 860 and 937 amino acids. The epitope specifically binds to antibodies from subjects diagnosed with endometriosis. The Repro-EN-1.0 and IB 1 proteins are useful for diagnosing endometriosis by detecting antibodies in immunoassays, and are used as markers for pathologic conditions e.g. breast, uterine or prostate cancer. Methods for detecting Repro-EN-1.0 or IB 1 polynucleotides or polypeptides are useful in the diagnosis of these cancers, monitoring their progress or treatment, and determining patient prognosis. Fragments of polynucleotides encoding Repro-EN-1.0 and IB 1 may be used as probes for detecting mRNA from cell types suspected to be cancerous, and as primers for amplifying sequences. Repro-EN-1.0 and IB 1 polypeptides and immunogenic fragments may be used as positive controls in diagnostic assays to detect antibodies that specifically bind the proteins from patient serum samples. The polypeptides are useful as immunogens for eliciting antibody production against epitopes of the protein, and as controls in diagnostic methods.

Sequence	860 AA
SQ	

99.78; Score 4514; DB 22; Length 800;

Query Match	2,546	Pred. No. 0;	Indels	0;	Gaps	0
Best Local Similarity	99.7%					
Matches	857;	Mismatches	1;			
Conservative	2;					

1 MSRGSTPHLLMDVKRKSLEDEE CATH...  
QY

Db 1 MSRGGSYPHLWDVRRKRLGLEDFSKLCKKALDQKRLT 120

61 DTGEYILSGSDDTKLVISNPYSRKVLITIKSGHNNALDNNH

DTGEYILSGSDDTKLVISNPYSRKVLTTIRSGHKANLE SANK EN C...

121 FYTNVEODAEINRQCFETCHYGTTEIMTVPNDPYTELSCGEDGI VNM D INKNT

123 FVTVNEODAEETNRCCQFTCHYGTYEIMTVPNDPYTFELSCGEDGIVKWE DININDC

101 DOZDOTTINCRATSAICPIPYLAVGCSDSVRIDRMLGTRATGNAGKGIOM 27

|||||TMCBPAATSVACPPPIPYLAVGCSDSVRIDRMLGTRATGNYAGRGITGM 24

DD 30-1-78  
ZNY  
FM JCRCMST CVS EDGO EILVSYSSDIYLLFDPKDDTARELKTPSAEEERE 30

[illegible]

Db 241 VAN LE CHIEU 36  
SPNYSIMORMSDMLSRWFEASEVA

0Y  
30L ELKQPEVANKLNSTN  
|||||  
|||BDCFOSPNVSI MORMSDMLSRWFEASEVA 36

Dd        301 ELRQPPVKRLNKGDMOZ-----  
                                PAFOFIOPSTSSITMSAQ 4:

QY 361 QSNRGGRSKPRGGISQSDICZLZ  
QY 361 QSNRGGRSKPRGGISQSDICZLZ

Db 361 QSNRGRGRSRPRGCTQSJLSTLPLVFCOZBZV...

421 AHSTSPTEPHSTPLSSPDSEQRKVS EASG.....\*

421 AHSTSSPTEPHSTPLSSPDSEQRQSVASGHHIHHQDNNVZ-----

081 YSTEGTTSIKLNFDEWSSIASSRGICSHCKSEGEESFVCCVX

481 YSTEGTTSTIKLNETDEWSSIASSRGIGSHCKSEGOEESFVPUSSVQFEBDDZ.....

241 RECEIVED KYOEGVSAENPVENHINTOSDKETAKPLDSNGERNDLNLDKSCGVFEEB.

|||||  
|||XOECVSAENPVNHINITQSDKFTAKPLDSNGERNDLDRSCGVPEESA

QY 601 SSKAKPEPTSDOTSEATNTNNTNPEPQFOTEGPSAHEETSTRDSDALDPTDSDDD 660  
 Db 601 SSKAKPEPTSDOTSEATNTNNTNPEPQFOTEGPSAHEETSTRDSDALDPTDSDDD 660  
 QY 661 PVLIGARYRAGPGRSARVARIOEFRRRKRKEKEEEDLTINIRPLVKMYKGRNSR 720  
 Db 661 PVLIGARYRAGPGRSARVARIOEFRRRKRKEKEEEDLTINIRPLVKMYKGRNSR 720  
 QY 721 TWIKANFWANFWNSGDCGHIFTWDRHTAHEMLLEADNHNVCLOPPEPDIASSG 780  
 Db 721 TWIKANFWANFWNSGDCGHIFTWDRHTAHEMLLEADNHNVCLOPPEPDIASSG 780  
 QY 781 IDYDIKISPLEESRIENRKLADDEVITRNEMLLEETRTITYPASPMLASLNHTRAD 840  
 Db 781 IDYDIKISPLEESRIENRKLADDEVITRNEMLLEETRTITYPASPMLASLNHTRAD 840  
 QY 841 RLEGDRSGSGQENENEDEE 860  
 Db 841 RLEGDRSGSGQENENEDEE 860

## RESULT 2

AAB60505  
 ID AAB60505 standard; Protein: 880 AA.

AC AAB60505;

DT 24-APR-2001 (first entry)

DE Human cell cycle and proliferation protein CCYPR-53, SEQ ID NO:53.

KM Cell cycle and proliferation protein: CCYPR; human; agonist;

KM antagonist; gene therapy; detection; gene therapy;

KM transgenic animal disease model; immune disorder;

KM developmental disorder; cell signalling disorder;

KM cell proliferative disorder; cancer; tumour; anaemia;

KM arteriosclerosis; asthma; allergy; diabetes mellitus;

KM menstrual cycle disorder; bacterial infection.

OS Homo sapiens.

XX WC0200107471-A2.

XX 01-FEB-2001.

XX 21-JUL-2000; 2000MO-US19948.

XX 21-JUL-1999; 99US-0145075.

XX 08-SEP-1999; 99US-0153129.

XX 10-NOV-1999; 99US-0164447.

XX (INCY-) INCYTE-GENOMICS INC.

XX Hillman JL, Lal P, Tang YT, Yue H, Au-Young J, Bandman O;

XX Azimtal Y, Yang J, Lu DAM, Baughn MR, Patterson C, Shah P;

XX MPI; 2001-11-27/72.

XX N-PSDB; AAF59642.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

CC compounds e.g., antibodies, oligonucleotides and proteins (receptors)  
 CC that specifically bind to CCYPR, and in drug screening methods to  
 CC identify compounds that modulate the activity of CCYPR. CCYPR  
 CC nucleotides can be used to generate transgenic animal models of human  
 CC disease, and can be used in gene therapy in target cells with genetic  
 CC abnormalities with respect to the expression of CCYPR for the  
 CC treatment or prevention of a disorder associated with CCYPR.  
 CC diseases which can be diagnosed, treated and prevented using CCYPR  
 CC proteins, nucleic acids, agonists or antagonists include immune,  
 CC developmental and cell signalling disorders, and cell proliferative  
 CC disorders including cancer. Specific examples of these disorders  
 CC include anaemia, epilepsy, arteriosclerosis, asthma, cancer, allergies,  
 CC diabetes mellitus, disorders of the menstrual cycle and infections  
 CC caused by bacteria.

Sequence 880 AA:

Query Match 99.4%; Score 4498; DB 22; Length 880;

Best Local Similarity 97.6%; Pred. No. 0;

Matches 859; Conservative 0; Mismatches 1; Indels 20; Gaps 1;

QY 1 MSRGSGYPLHMDVYKRSIGLEDPKRLSRVYGRREFIQRLKLEATLVNHDCCVNTICWN 60  
 Db 1 MSRGSGYPLHMDVYKRSIGLEDPKRLSRVYGRREFIQRLKLEATLVNHDCCVNTICWN 60  
 QY 61 DTGEYILSGDDTKLYISNPYSRKVLTTRSGHRANIFSAKFLPCTNDKQIYSGSDGYI 120  
 Db 61 DTGEYILSGDDTKLYISNPYSRKVLTTRSGHRANIFSAKFLPCTNDKQIYSGSDGYI 120  
 QY 121 FYTNVEDAETNRQOCFTCHVGTVEIMTVPNDPYFLSCGEGTVRWFDRITKSTCKE 180  
 Db 121 FYTNVEDAETNRQOCFTCHVGTVEIMTVPNDPYFLSCGEGTVRWFDRITKSTCKE 180  
 QY 121 FYTNVEDAETNRQOCFTCHVGTVEIMTVPNDPYFLSCGEGTVRWFDRITKSTCKE 180  
 Db 121 FYTNVEDAETNRQOCFTCHVGTVEIMTVPNDPYFLSCGEGTVRWFDRITKSTCKE 180  
 QY 181 DCKDDILNCRANASVAICPPIPYLLAVGCDSSVRYIDRMGLTRATNAGAGTTGM 240  
 Db 181 DCKDDILNCRANASVAICPPIPYLLAVGCDSSVRYIDRMGLTRATNAGAGTTGM 240  
 QY 241 VARETPSHLNKNSCRVTSICYSDEQELIVSYSDYIYLFDPKDTARELTPKPSAEERE 300  
 Db 241 VARETPSHLNKNSCRVTSICYSDEQELIVSYSDYIYLFDPKDTARELTPKPSAEERE 300  
 QY 301 ELROPVYKRLRLRGDSDTGPAPRESERERDEQSPNVSILQMRSDMLSRFEEASAVA 360  
 Db 301 ELROPVYKRLRLRGDSDTGPAPRESERERDEQSPNVSILQMRSDMLSRFEEASAVA 360  
 QY 361 QSNRGKRSRPRGTSOSDISTLPYVSSPDLEVESETAMEVDTPABOFLQSTSTMSAQ 420  
 Db 361 QSNRGKRSRPRGTSOSDISTLPYVSSPDLEVESETAMEVDTPABOFLQSTSTMSAQ 420  
 QY 421 AHSTSSPTSPHSTPLSSPDSEOROSVEASGHTHHQS-----D 460  
 Db 421 AHSTSSPTSPHSTPLSSPDSEOROSVEASGHTHHQS-----D 460  
 QY 461 NNNKLSKPKGTGEPVLSLHSTGTTSTIKLNTDWMSSIASRSGISGHCSEGE 520  
 Db 461 NNNKLSKPKGTGEPVLSLHSTGTTSTIKLNTDWMSSIASRSGISGHCSEGE 520  
 QY 521 SFVPOSSVQPEGSEETAPRESSDVTYKQEGVSAENPVNHNITQSDFTAKPLDSN 580  
 Db 521 SFVPOSSVQPEGSEETAPRESSDVTYKQEGVSAENPVNHNITQSDFTAKPLDSN 580  
 QY 581 SGERNDLNDRCGVPESASSEKAKPEPSTDSQSTESATNNTNPEPQFOTEGTATGSA 640  
 Db 581 SGERNDLNDRCGVPESASSEKAKPEPSTDSQSTESATNNTNPEPQFOTEGTATGSA 640  
 QY 641 HEETSTRDSDALDPTDSDDDVLLPGARYRAGPGRSARVARIOEFRRRKRKEEEMELD 700  
 Db 641 HEETSTRDSDALDPTDSDDDVLLPGARYRAGPGRSARVARIOEFRRRKRKEEEMELD 700  
 QY 701 TLTNRPLVKKVYVGHNSRTMIKANFWANFWNSGDCGHIFTWDRHTAHEMLLEAD 760  
 Db 701 TLTNRPLVKKVYVGHNSRTMIKANFWANFWNSGDCGHIFTWDRHTAHEMLLEAD 760

QY 761 NHVNCLOPHPPDPIASSGIDYDIKIMSPLEESRIFFNRKLADEVITRNELMEETRNIT 820  
 DB 781 NHVNCLOPHPPDPIASSGIDYDIKIMSPLEESRIFFNRKLADEVITRNELMEETRNIT 840  
 QY 821 TVPASFMLRLASLNHIRADRLGDRSGSGQENENDEE 860  
 DB 841 TVPASFMLRLASLNHIRADRLGDRSGSGQENENDEE 880

## RESULT 3

AAB69557  
 ID AAB69557 standard; Protein; 937 AA.

AC AAB69557;

DT 27-APR-2001 (first entry)

DE Human IB1 protein.

KW Human: Repro-EN-1.0; IB1: cytostatic; breast cancer; uterine cancer;

KM prostate cancer; epitope.

OS Homo sapiens.

PN MO200107616-A1.

PD 01-FEB-2001.

PF 10-MAR-2000; 2000MO-US06742.

PR 22-JUL-1999; 99US-0359084.

PR 30-JUL-1999; 99MO-US17284.

PR 23-NOV-1999; 99US-0447399.

PI (DIAG-) DIAGNOSTIC PROD CORP.

PI El Sham1 AS; Menon SN; French CK;

DR WPI; 2001-182795/18.

DR N-PSDB; AAF58655.

PT New autoantigens Repro-EN-1.0 and IB 1 polypeptides and polynucleotides

PT conditions such as breast, uterine or prostate cancer

PS Claim 24; Page 29-34; 119pp; English.

CC The present sequence is given in a specification relating to recombinant  
 CC polynucleotides comprising nucleotide sequences encoding a polypeptide  
 CC epitope of at least 5 amino acids of Repro-EN-1.0 or of IB 1 having a  
 CC fully defined sequence of 860 and 937 amino acids. The epitope  
 CC specifically binds to antibodies from subjects diagnosed with  
 CC endometriosis. The Repro-EN-1.0 and IB 1 proteins are useful for  
 CC diagnosing endometriosis by detecting antibodies in immunoassays, and  
 CC are used as markers for pathologic conditions e.g. breast, uterine or  
 CC prostate cancer. Methods for detecting Repro-EN-1.0 or IB 1  
 CC polynucleotides or polypeptides are useful in the diagnosis of these  
 CC cancers, monitoring their progress or treatment, and determining patient  
 CC prognosis. Fragments of polynucleotides encoding Repro-EN-1.0 and IB 1  
 CC may be used as probes for detecting mRNA from cell types suspected to  
 CC be cancerous, and as primers for amplifying sequences. Repro-EN-1.0 and  
 CC IB 1 polypeptides and immunogenic fragments may be used as positive  
 CC controls in diagnostic assays to detect antibodies that specifically  
 CC bind the proteins from patient serum samples. The polypeptides are  
 CC useful as immunogens for eliciting antibody production against epitopes  
 CC of the protein, and as controls in diagnostic methods.

CC Sequence 937 AA;

Query Match 98.6%; SCORE 4465.5; DB 22; Length 937;

Best local similarity 91.5%; Pred. No. 3.8e-318;

Matches 857; Conservative 2; Mismatches 1; Indels 77; Gaps 1;

QY 1 MSRGGSYPHLLMDVNRKSLGLIEDPSRLRSRYLGRREFIQRKLKLEATLVHDCVNTICWN 60  
 DB 1 MSRGGSYPHLLMDVNRKSLGLIEDPSRLRSRYLGRREFIQRKLKLEATLVHDCVNTICWN 60  
 QY 61 DTGEYIISGDDPTKLYISNYSRKVLTITISGRANIFSAKFLPCTNDKQIVSGSGGY 120  
 DB 61 DTGEYIISGDDPTKLYISNYSRKVLTITISGRANIFSAKFLPCTNDKQIVSGSGGY 120  
 QY 121 FYTNVEDAEFTNQCQPTCHGYTYEIMTVPNPDYTFILSCGEDGYVMPDRIKTSCKE 180  
 DB 121 FYTNVEDAEFTNQCQPTCHGYTYEIMTVPNPDYTFILSCGEDGYVMPDRIKTSCKE 180  
 QY 181 DCKDDILINCRAATVAICPPIPYIYAVGSDSSVRIYDRMLGTATGNYAGRTTGM 240  
 DB 181 DCKDDILINCRAATVAICPPIPYIYAVGSDSSVRIYDRMLGTATGNYAGRTTGM 240  
 QY 241 VAFIPSHLNKSCRYSLCYSEDGQELIVSYSSDYLYLPDPKDTARELKTPSAERRE 300  
 DB 241 VAFIPSHLNKSCRYSLCYSEDGQELIVSYSSDYLYLPDPKDTARELKTPSAERRE 300  
 QY 301 ELROPVYKRLRLRGDWSDTGPRAPESBEREDGQSPNVSLMQRMSDLRMFEASEVA 360  
 DB 301 ELROPVYKRLRLRGDWSDTGPRAPESBEREDGQSPNVSLMQRMSDLRMFEASEVA 360  
 QY 361 QSNRGGRSRPRGTSQSDISTLPTVPSPPLEYSFETAMEVDTPAEQFLDPSTSTSAQ 420  
 DB 361 QSNRGGRSRPRGTSQSDISTLPTVPSPPLEYSFETAMEVDTPAEQFLDPSTSTSAQ 420  
 QY 421 AHSTSSPESHSTPILSSPDSSEOROSVEASGHTTHOS----- 459  
 DB 421 AHSTSSPESHSTPILSSPDSSEOROSVEASGHTTHOS----- 459  
 QY 460 -----DNNN 463  
 DB 460 -----DNNN 463  
 QY 481 KAEQROEELAAHTQQQPTSDQSHEGSSQDPHASPSSSVYVKNKQLGMSLDEQODNNN 540  
 DB 481 KAEQROEELAAHTQQQPTSDQSHEGSSQDPHASPSSSVYVKNKQLGMSLDEQODNNN 540  
 QY 464 EKLSPKRGTPVLSLHSTEGTTSTIKLNFTEWSSILASSRGIGSHCKSEGOEESFV 523  
 DB 464 EKLSPKRGTPVLSLHSTEGTTSTIKLNFTEWSSILASSRGIGSHCKSEGOEESFV 523  
 QY 541 EKLSPKRGTPVLSLHSTEGTTSTIKLNFTEWSSILASSRGIGSHCKSEGOEESFV 600  
 DB 541 EKLSPKRGTPVLSLHSTEGTTSTIKLNFTEWSSILASSRGIGSHCKSEGOEESFV 600  
 QY 524 POSSVQPPGDSSETKAPESESEDVTKYQGVSAENPVNHNITQSDFTAKPLDSSNGE 583  
 DB 524 POSSVQPPGDSSETKAPESESEDVTKYQGVSAENPVNHNITQSDFTAKPLDSSNGE 583  
 QY 601 POSSVQPPGDSSETKAPESESEDVTKYQGVSAENPVNHNITQSDFTAKPLDSSNGE 660  
 DB 601 POSSVQPPGDSSETKAPESESEDVTKYQGVSAENPVNHNITQSDFTAKPLDSSNGE 660  
 QY 584 RNDLMDRSCGVPESASSEKAKEPETSDQSTESATNENTNPNPQOTATGSAHEE 643  
 DB 584 RNDLMDRSCGVPESASSEKAKEPETSDQSTESATNENTNPNPQOTATGSAHEE 643  
 QY 661 RNDLMDRSCGVPESASSEKAKEPETSDQSTESATNENTNPNPQOTATGSAHEE 720  
 DB 661 RNDLMDRSCGVPESASSEKAKEPETSDQSTESATNENTNPNPQOTATGSAHEE 720  
 QY 644 TSTRSALODDDDDDDVLLPGARYRAGPGDRSAVARIOEFFRRRERKMEELDTLN 703  
 DB 644 TSTRSALODDDDDDDVLLPGARYRAGPGDRSAVARIOEFFRRRERKMEELDTLN 703  
 QY 721 TSTRSALODDDDDDDVLLPGARYRAGPGDRSAVARIOEFFRRRERKMEELDTLN 780  
 DB 721 TSTRSALODDDDDDDVLLPGARYRAGPGDRSAVARIOEFFRRRERKMEELDTLN 780  
 QY 704 IRRPLVKKVYKGRHSRTMIKEANFWANFVWSSDCGHIFIMDRHTAEHLMLLEADNHV 763  
 DB 704 IRRPLVKKVYKGRHSRTMIKEANFWANFVWSSDCGHIFIMDRHTAEHLMLLEADNHV 763  
 QY 764 VNCLOPHPPDPIASSGIDYDIKIMSPLEESRIFFNRKLADEVITRNELMEETRNITVP 823  
 DB 764 VNCLOPHPPDPIASSGIDYDIKIMSPLEESRIFFNRKLADEVITRNELMEETRNITVP 823  
 QY 841 VNCLOPHPPDPIASSGIDYDIKIMSPLEESRIFFNRKLADEVITRNELMEETRNITVP 900  
 DB 841 VNCLOPHPPDPIASSGIDYDIKIMSPLEESRIFFNRKLADEVITRNELMEETRNITVP 900  
 QY 824 ASFMLRLASLNHIRADRLGDRSGSGQENENDEE 860  
 DB 901 ASFMLRLASLNHIRADRLGDRSGSGQENENDEE 937

## RESULT 4

ABB05707  
 ID ABB05707 standard; Protein; 482 AA.

AC ABB05707;

DT 30-APR-2002 (first entry)

XX



DE Human signal transduction protein clone test\_11c22.  
 XX  
 KW Human; foetal brain; foetal kidney; melanoma; testis; amygdala;  
 KW gene therapy; chromosome 1q23.2-24.3.  
 XX  
 OS Hemo sapiens.  
 XX  
 PY WO200198454-A2.  
 XX  
 PK 27-DEC-2001.  
 XX  
 PF 25-APR-2001; 2001WO-1B02050.  
 XX  
 PR 25-APR-2000; 2000US-199380P.  
 XX  
 PR (GEHU-) GERMAN HUMAN GENOME PROJECT.  
 XX  
 PI Wiemann S;  
 XX  
 DR WPI: 2002-055860/07.  
 DR N-PSDB: ABA93744.  
 XX  
 PT Human cDNA sequences and clones derived from human fetal brain, fetal  
 PT kidney, melanoma, testis and amygdala cDNA libraries, useful in genetic  
 PT screening and therapy -  
 XX  
 PS Claim 1; Page 287; 611pp; English.  
 XX  
 CC The present invention describes assemblages and computer readable media  
 CC comprising novel human cDNA sequences and clones derived from human  
 CC foetal brain, foetal kidney, melanoma, testis and amygdala cDNA  
 CC libraries. ABA93702 to ABA93766 represent human cDNA sequences from the  
 CC present invention which encode the proteins given in ABB05662 to  
 CC ABB05729. The human cDNA sequences and clones can be used in gene  
 CC therapy. The clones may be used in a variety of applications, for  
 CC example they may be used in profiling assays, for providing large arrays  
 CC of human genetic material for implementing large-scale screening  
 CC strategies and for treating diseases via gene therapy procedures.  
 CC  
 XX  
 SO Sequence 482 AA;  
 Query Match 52.8%; Score 2390; DB 23; Length 482;  
 Best Local Similarity 95.9%; Pred. No. 1.5e-166;  
 Matches 462; Conservative 0; Mismatches 0; Indels 20; Gaps 1;  
 QY 399 MEVTPAEOFLQFSTSTMSQAHSSTSPSTPLSSPDSEQROSVASGHHTHQ 458  
 DB 1 MEVTPAEOFLQFSTSTMSQAHSSTSPSTPLSSPDSEQROSVASGHHTHQ 60  
 QY 459 S-----DNNNEKLSPPKPGGEPVLSHYSTEGTSTIKINFTDE 498  
 DB 61 SDSPSSVYNKQSGMSLDEQDDNNNEKLSPPKPGGEPVLSHYSTEGTSTIKINFTDE 120  
 QY 499 WSSIASSSRGIGSHCKSEGESEFVPOSSVOPPGDSETKAPESSEEDVTYQEGVSAEN 558  
 DB 121 WSSIASSSRGIGSHCKSEGESEFVPOSSVOPPGDSETKAPESSEEDVTYQEGVSAEN 180  
 QY 559 PVEHHNITQSDKRTAPLDSNNGERDNLNDRSCGVPESASSEKAKEPTSDQSTES 618  
 DB 181 PVEHHNITQSDKRTAPLDSNNGERDNLNDRSCGVPESASSEKAKEPTSDQSTES 240  
 QY 619 ATNNNNNPPEPQTEATGSAHEETSTRSALODTDDSDDDPVLIGARARAPGRRS 678  
 DB 241 ATNNNNNPPEPQTEATGSAHEETSTRSALODTDDSDDDPVLIGARARAPGRRS 300  
 QY 679 AVARIOEFFRRRERKEMEELDTLIRPLVKMYKGRHSRTMIKEANFGANFVAGS 738  
 DB 301 AVARIOEFFRRRERKEMEELDTLIRPLVKMYKGRHSRTMIKEANFGANFVAGS 360  
 QY 739 DCGHIFIMDHTAEHMLLEADNVVNCLOPHRDPILASSGIDYDIKINSPLSESRIFN 798  
 DB 361 DCGHIFIMDHTAEHMLLEADNVVNCLOPHRDPILASSGIDYDIKINSPLSESRIFN 420

QY 799 RKLADEVITRNELMEETRNITIVPASFMLMLASINHIRADRLREGDSGCGENED 858  
 DB 421 RKLADEVITRNELMEETRNITIVPASFMLMLASINHIRADRLREGDSGCGENED 480  
 QY 859 EE 860  
 DB 481 EE 482  
 RESULT 5  
 AAU19599  
 ID AAU19599 standard; Protein; 301 AA.  
 XX  
 AC AAU19599;  
 XX  
 DT 06-DEC-2001 (first entry)  
 XX  
 DE Human diagnostic and therapeutic polypeptide (DTRP) #185.  
 XX  
 KW Human; receptor; diagnostic; therapeutic; gene therapy; vaccine;  
 KW cell proliferative disorder; Crohn's disease; lymphoma; leukaemia;  
 KW acquired immune deficiency syndrome; AIDS; autoimmune disorder;  
 KW respiratory disorder.  
 OS Hemo sapiens.  
 XX  
 PN WO200162927-A2.  
 XX  
 PD 30-AUG-2001.  
 XX  
 PF 21-FEB-2001; 2001WO-US06059.  
 XX  
 PR 24-FEB-2000; 2000US-0184693.  
 PR 24-FEB-2000; 2000US-0184697.  
 PR 24-FEB-2000; 2000US-0184698.  
 PR 24-FEB-2000; 2000US-0184768.  
 PR 24-FEB-2000; 2000US-0184769.  
 PR 24-FEB-2000; 2000US-0184770.  
 PR 24-FEB-2000; 2000US-0184771.  
 PR 24-FEB-2000; 2000US-0184772.  
 PR 24-FEB-2000; 2000US-0184773.  
 PR 24-FEB-2000; 2000US-0184776.  
 PR 24-FEB-2000; 2000US-0184777.  
 PR 24-FEB-2000; 2000US-0184797.  
 PR 24-FEB-2000; 2000US-0184813.  
 PR 24-FEB-2000; 2000US-0184837.  
 PR 24-FEB-2000; 2000US-0184841.  
 PR 24-FEB-2000; 2000US-0185213.  
 PR 12-MAY-2000; 2000US-0185216.  
 PR 12-MAY-2000; 2000US-0203785.  
 PR 15-MAY-2000; 2000US-0204226.  
 PR 16-MAY-2000; 2000US-0204525.  
 PR 16-MAY-2000; 2000US-0204821.  
 PR 16-MAY-2000; 2000US-0204908.  
 PR 16-MAY-2000; 2000US-0205232.  
 PR 17-MAY-2000; 2000US-0204815.  
 PR 17-MAY-2000; 2000US-0204863.  
 PR 17-MAY-2000; 2000US-0205221.  
 PR 17-MAY-2000; 2000US-0205285.  
 PR 17-MAY-2000; 2000US-0205286.  
 PR 17-MAY-2000; 2000US-0205287.  
 PR 17-MAY-2000; 2000US-0205323.  
 PR 17-MAY-2000; 2000US-0205324.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 PI Panzer SF, Spito PA, Banville SC, Shah P, Chalup MS, Chang SC;  
 PI Chen A, D'sa SA, Amshay S, Dahl CR, Danilels SE;  
 PI Difour GE, Flores V, Fong WT, Greenwalt LB, Hillman JL, Jones AL;  
 PI Liu TF, Roseberry AM, Rosen BH, Russo FD, Stockdreher TK, Datfo A;  
 PI Wright RJ, Yap PE, Yu JY, Bradley DL, Bratcher SK, Chen W;  
 PI Cohen HU, Hodgson DM, Lincoln SE, Jackson SJ;

XX WPI: 2001-502867/55.  
DR N-PSDB: AAS31170.

PT Polynucleotides encoding diagnostic and therapeutic proteins, e.g.  
PT enzymes, hormones and receptors, useful in diagnostics and therapeutics

PS Claim 27; Page 507; 522pp; English.

XX The invention relates to polynucleotides (I) encoding diagnostic and  
CC therapeutic (DTRP) polypeptides (II), which include e.g. enzymes,  
CC and proteins involved in growth and development and receptors. (I) and  
CC (II) may be used in the prevention, diagnosis and treatment of diseases  
CC associated with inappropriate DTRP expression. For example, (I) and  
CC (II) may be used to treat disorders associated with decreased polypeptide  
CC expression by rectifying mutations or deletions in a patient's genome,  
CC that affect the activity of the DTRPs, by expressing inactive proteins  
CC or supplementing the patient's own production of them. (I) and (II)  
CC may be used to treat diseases, for example, cell proliferative disorder,  
CC Crohn's disease, acquired immune deficiency syndrome (AIDS), lymphoma,  
CC leukaemia, autoimmune disorders, and respiratory disorders. Additionally,  
CC (I) may be used to produce the DTRPs, by inserting the nucleic acids  
CC into a host cell and culturing the cell to express the protein. (I) and  
CC its complementary sequences may also be used as DNA probes in diagnostic  
CC assays to detect and quantitate the presence of similar nucleic acids in  
CC samples, and therefore which patients may be in need of restorative  
CC therapy. (II) may also be used as antigens in the production of  
CC antibodies against DTRPs and in assays to identify modulators of DTRP  
CC expression and activity. The anti-DTRP antibodies and antagonists may  
CC also be used to down regulate expression and activity. The anti-DTRP  
CC antibodies may also be used as diagnostic agents for detecting the  
CC presence of DTRPs in samples (e.g. by enzyme linked immunosorbent  
CC assay (ELISA)). AAU19415-AAU19625 represent human diagnostic and  
CC therapeutic (DTRP) polypeptides of the invention.

XX Sequence 301 AA:

Query Match 32.2%; Score 1459; DB 22; Length 301;  
Best Local Similarity 95.0%; Pred. No. 1.3e-98;  
Matches 283; Conservative 0; Mismatches 1; Indels 14; Gaps 1;

QY 459 SDNNKSLSPKRGTEPVLSTHYSTEGTSTIKNTFDEWSIASSSSGTSHCKSEQ 518  
DB 4 SDNNKSLSPKRGTEPVLSTHYSTEGTSTIKNTFDEWSIASSSSGTSHCKSEQ 63  
QY 519 ESEFVPOSSVOPPEGDSSTKAPESSEEDVTKEQGVSAENPVENHINITQSDKFTAKPLD 578  
DB 64 ESEFVPOSSVOPPEGDSSTKAPESSEEDVTKEQGVSAENPVENHINITQSDKFTAKPLD 123  
QY 579 SNSGERNDLNDRCGVPESASSEKAKEPETSDDTSTESATNENNTNPEPOFOTBANGP 638  
DB 124 SNSGERNDLNDRCGVPESASSEKAKEPETSDDTSTESATNENNTNPEPOFOTBANGP 183  
QY 639 SAHEETSPRDSALODTDDSDDDPVLIPGARVRAGPGD-----RRAVARIQ 684  
DB 184 SAHEETSPRDSALODTDDSDDDPVLIPGARVRAGPGDRNINIGTTIGDRIMRRAVARIQ 243  
QY 685 EEFRRRRKKEKEEELDTNIRRLPVKMYKGRNSRTMIKEANFNGANFVMSGSDCGH 742  
DB 244 EEFRRRRKKEKEEELDTNIRRLPVKMYKGRNSRTMIKEANFNGANFVMSGSDCGH 301

RESULT 6  
AAB92537  
ID AAB92537 standard; Protein: 163 AA.

AC AAB92537;

XX 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO:10703.  
XX

KW Human; primer: detection; diagnosis; antisense therapy; gene therapy.  
XX  
XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI) HELIX-RBS-INST.

XX Ota T, Isogai S, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI: 2001-318749/34.

PT Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX  
XX  
XX Claim 8; SEQ ID 10703; 2537pp + CD ROM; English.

CC The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.

XX Sequence 163 AA:

Query Match 18.7%; Score 847; DB 22; Length 163;  
Best Local Similarity 98.2%; Pred. No. 3.6e-54;  
Matches 162; Conservative 0; Mismatches 1; Indels 2; Gaps 1;

QY 696 MEELDTNIRRLPVKMYKGRNSRTMIKEANFNGANFVMSGSDCGHIFIMDRHAHELM 755  
DB 1 MEELDTNIRRLPVKMYKGRNSRTMIKEANFNGANFVMSGSDCGHIFIMDRHAHELM 60

QY 756 LLEADNHVNCIQPHFDPYLLASGIDYDIKIKWSPLEESRIRNRKADAVYIRNEMLEEF 815  
DB 61 LLEADNHVNCIQPHFDPYLLASGIDYDIKIKWSPLEESRIRNRKADAVYIRNEMLEEF 118

QY 816 TRNTIVPASFMRLMLASLNHTRADRLGDRSGSGQENENEDDE 860  
DB 119 TRNTIVPASFMRLMLASLNHTRADRLGDRSGSGQENENEDDE 163

RESULT 7

AAB27984  
ID AAB27984 standard; Protein; 129 AA.  
XX  
AC AAB27984;  
XX  
DT 02-FEB-2001 (first entry)  
XX  
DE Human secreted protein BLAST search protein SEQ ID NO: 138.  
XX  
KM Cytostatic; immunosuppressive; neutrotropic; neuroprotective; antiviral;  
KM antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;  
KM vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KM cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KM neurological disease; infection; human; secreted protein.  
XX  
OS Homo sapiens.  
XX  
PY WO200055171-A1.  
XX  
HD 21-SEP-2000.  
XX  
PF 09-MAR-2000; 2000WO-US06043.  
XX  
PR 12-MAR-1999; 99US-0124146.  
PR 23-NOV-1999; 99US-0167061.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
PI Rosen CA, Ruben SM, Komatsoulis G;  
XX  
XX WPI; 2000-638174/61.  
XX  
DR Isolated nucleic acid molecule encoding a human secreted protein is  
XX used in preventing, treating or ameliorating a medical condition -  
XX  
XX Disclosure; Page 400; 438pp; English.  
XX  
CC The invention relates to the isolation of genes AAC59049-C59098 encoding  
CC the human secreted proteins AAB27907-B27956. This sequence represents a  
CC fragment of the protein encoded by the gene given in the descriptor  
CC line. The sequence is used as a query sequence for doing BLASTX searches  
CC to determine homologous sequence to the protein. The genes and proteins  
CC are useful for preventing, ameliorating or treating medical conditions,  
CC e.g. by protein or gene therapy. The genes are isolated from a range of  
CC human tissues disclosed in the specification. The nucleic acids,  
CC proteins, antibodies and (ant)agonists are useful in the diagnosis,  
CC treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer,  
CC and other cancers of the adrenal gland, bone, bone marrow, breast,  
CC gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders  
CC e.g. Addison's disease, allergies, autoimmune haemolytic anaemia,  
CC autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple  
CC sclerosis, rheumatoid arthritis and ulcerative colitis; (c)  
CC cardiovascular disorders such as myocardial ischaemias; (d) wound  
CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and  
CC (f) infectious diseases such as viral, bacterial, fungal and parasitic  
CC infections.  
XX  
SQ Sequence 129 AA;  
XX  
Query Match 15.2%; Score 688; DB 21; Length 129;  
Best Local Similarity 100.0%; Pred. No. 1.1e-42;  
Matches 129; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 121 SLNHIRADR 129  
RESULT 8  
ID AAB27926  
XX  
AC AAB27926;  
XX  
DT 02-FEB-2001 (first entry)  
XX  
DE Human secreted protein SEQ ID NO: 80.  
XX  
KM Cytostatic; immunosuppressive; neutrotropic; neuroprotective; antiviral;  
KM antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;  
KM vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KM cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KM neurological disease; infection; human; secreted protein.  
XX  
OS Homo sapiens.  
XX  
PY WO200055171-A1.  
XX  
HD 21-SEP-2000.  
XX  
PF 09-MAR-2000; 2000WO-US06043.  
XX  
PR 12-MAR-1999; 99US-0124146.  
PR 23-NOV-1999; 99US-0167061.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
PI Rosen CA, Ruben SM, Komatsoulis G;  
XX  
XX WPI; 2000-638174/61.  
XX  
DR N-PSDB; AAC59068.  
XX  
XX Isolated nucleic acid molecule encoding a human secreted protein is  
XX used in preventing, treating or ameliorating a medical condition -  
XX  
XX Claim 11; Page 371; 438pp; English.  
XX  
CC Sequences AAB27907-B27956 represent the amino acid sequences of 50  
CC human secreted proteins encoded by the genes AAC59049-C59098. The genes  
CC and proteins are useful for preventing, ameliorating or treating medical  
CC conditions, e.g. by protein or gene therapy. The genes are isolated from  
CC a range of human tissues disclosed in the specification. The nucleic  
CC acids, proteins, antibodies and (ant)agonists are useful in the  
CC diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
CC ovarian cancer, and other cancers of the adrenal gland, bone, bone  
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,  
CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.  
XX  
SQ Sequence 127 AA;  
XX  
Query Match 14.7%; Score 664; DB 21; Length 127;  
Best Local Similarity 100.0%; Pred. No. 6.4e-41;  
Matches 126; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 855 ENEDEE 860  
 Db 121 ENEDEE 126

RESULT 9  
 ABB29956  
 ID ABB29956 standard; Peptide: 110 AA.

AC ABB29956;

DT 01-FEB-2002 (first entry)

DE Peptide #2607 encoded by breast cell single exon nucleic acid probe.

KW Human; microarray; single exon probe; gene expression; breast;  
 disease; cancer.

OS Homo sapiens.

PN WO200157271-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00662.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-496933/54.

PT New spatially-addressable set of single exon nucleic acid probes,  
 useful for measuring gene expression in sample derived from human

PS breast, comprises number of single exon nucleic acid probes

XX Claim 27; SEQ ID NO 12924; 327bp + sequence listing; English.

XX The invention relates to a spatially-addressable set of single exon  
 nucleic acid probes for measuring gene expression in a sample derived

CC from human breast and BT 474 cells. The method involves contacting  
 the probes with a collection of detectably labeled nucleic acids

CC derived from mRNA of human breast, and then measuring the label  
 bound to each probe of the microarray. The probes are useful for

CC verifying the expression of regions of genomic DNA predicted to  
 encode proteins. They are useful for gene discovery, and for

CC determining predisposition and/or prognosing breast disease. Gene  
 expression analysis is useful for assessing the toxicity of chemical

CC agents on cells. The microarray of this invention presents a far greater  
 diversity of probes for measuring gene expression, with far less bias

CC than expressed sequence tag microarrays. The method is suitable for  
 rapid production of functional information from genomic sequence. The

CC present sequence is a peptide encoded by a single exon nucleic acid  
 probe of the invention.

CC Note: The sequence data for this patent did not form part of the  
 printed specification, but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 110 AA;  
 SQ

Db 1 QSDKFTAKPLDLSNGERNDLNDRSCGVPESASSEKAKEPETSQSTESATNENNTNP 60

QY 628 EPQFTEATGPSAHEETSTRDSALQDTDDSDPVLIGARYRAGPGR 676

Db 61 EPQFTEATGPSAHEETSTRDSALQDTDDSDPVLIGARYRAGPGR 109

RESULT 10  
 ABB20555  
 ID ABB20555 standard; Protein: 110 AA.

AC ABB20555;

DT 23-JAN-2002 (first entry)

DE Protein #2554 encoded by probe for measuring heart cell gene expression.

KW Human; gene expression; heart; microarray; vascular system;  
 cardiovascular disease; hypertension; cardiac arrhythmia;

KW congenital heart disease.

OS Homo sapiens.

PN WO200157274-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00666.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-488899/53.

PT Single exon nucleic acid probes for analyzing gene expression in human  
 hearts

PS Claim 15; SEQ ID NO 22325; 530bp; English.

XX The present invention relates to single exon nucleic acid probes for  
 measuring human gene expression in a sample derived from human heart (see

CC ABA21535-ABA41305). The present sequence is a protein encoded by one such  
 CC probe. The probes may be used for predicting, measuring and displaying

CC gene expression in samples derived from the human heart via microarray.  
 By measuring gene expression, the probes are useful for predicting,

CC diagnosing, grading, staging, monitoring and prognosing diseases of the  
 CC human heart and vascular system e.g. cardiovascular disease,

CC hypertension, cardiac arrhythmias and congenital heart disease.  
 CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 110 AA;  
 SQ

Query Match 12.6%; Score 569; DB 22; Length 110;  
 Best Local Similarity 100.0%; Pred. No. 4.7e-34;

Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 628 EPQFTEATGPSAHEETSTRDSALQDTDDSDPVLIGARYRAGPGR 676

DB 61 EPOFOTGATGSAHEETSTRDSALQDTDDSDPVLIPGARYRAGPDR 109

## RESULT 11

AAM55960  
ID AAM55960 standard; Protein; 110 AA.

AC AAM55960;

DT 05-NOV-2001 (first entry)

DE Human brain expressed single exon probe encoded protein SEQ ID NO: 28065.

KW Human; brain expressed exon; gene expression analysis; probe;  
microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
epilepsy; cancer.

OS Homo sapiens.

PN WO200157275-A2.

PD 09-AUG-2001.

PE 30-JAN-2001; 2001WO-US00667.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAR-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0633366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI: 2001-483446/52.

PT Single exon nucleic acid probes for analyzing gene expression in human  
brains -

PS Example 4; SEQ ID NO: 28065; 650bp + Sequence Listing; English.

CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC brain. They can be used to measure gene expression in brain cell samples,  
CC which may enable the diagnosis and improved treatment of nervous system  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is a protein encoded by one of  
CC the probes of the invention.

XX Sequence 110 AA;

Query Match 12.6%; Score 569; DB 22; Length 110;  
Best Local Similarity 100.0%; Pred. No. 4.7e-34;

Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 568 QSDKFTAKPLDLSNGERNDLNDRCGVPESASSEKAKEPTSDQSTESATNENNTNP 627  
DB 1 QSDKFTAKPLDLSNGERNDLNDRCGVPESASSEKAKEPTSDQSTESATNENNTNP 60

QY 628 EPOFOTGATGSAHEETSTRDSALQDTDDSDPVLIPGARYRAGPDR 676  
DB 61 EPOFOTGATGSAHEETSTRDSALQDTDDSDPVLIPGARYRAGPDR 109

## RESULT 12

AAM68328

ID AAM68328 standard; Protein; 110 AA.

XX AAM68328;

XX 06-NOV-2001 (first entry)

DE Human bone marrow expressed probe encoded protein SEQ ID NO: 28634.

KW Human; bone marrow expressed exon; gene expression analysis; probe;  
microarray; cancer; leukemia; lymphoma; myeloma.

OS Homo sapiens.

PN WO200157276-A2.

PD 09-AUG-2001.

PE 30-JAN-2001; 2001WO-US00668.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAR-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0633366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI: 2001-488900/53.

PT Human genome-derived single exon nucleic acid probes useful for  
analyzing gene expression in human bone marrow -

PS Example 4; SEQ ID NO: 28634; 658bp + Sequence Listing; English.

CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC bone marrow. They can be used to measure gene expression in bone marrow  
CC samples, which may enable the improved diagnosis and treatment of cancers  
CC such as lymphoma, leukemia and myeloma. The present sequence is a  
CC protein encoded by one of the probes of the invention.

XX Sequence 110 AA;

Query Match 12.6%; Score 569; DB 22; Length 110;  
Best Local Similarity 100.0%; Pred. No. 4.7e-34;

Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 568 QSDKFTAKPLDLSNGERNDLNDRCGVPESASSEKAKEPTSDQSTESATNENNTNP 627  
DB 1 QSDKFTAKPLDLSNGERNDLNDRCGVPESASSEKAKEPTSDQSTESATNENNTNP 60

QY 628 EPOFOTGATGSAHEETSTRDSALQDTDDSDPVLIPGARYRAGPDR 676  
DB 61 EPOFOTGATGSAHEETSTRDSALQDTDDSDPVLIPGARYRAGPDR 109

## RESULT 13

AAM16144

ID AAM16144 standard; Protein; 110 AA.

AC AAM16144;

DT 12-OCT-2001 (first entry)

DE Peptide #2578 encoded by probe for measuring cervical gene expression.

KW Probe; human; microarray; gene expression; cervical epithelial cell;

OS Homo sapiens.

PN WO200157278-A2.

XX 09-AUG-2001.  
 PD 30-JAN-2001; 2001WO-US00670.  
 PF  
 XX 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 DR WPI; 2001-488901/53.  
 XX  
 PT Human genome-derived single exon nucleic acid probes useful for  
 PS analyzing gene expression in human cervical epithelial cells -  
 CC Claim 27; SEQ ID NO 20970; 487pp; English.  
 XX  
 CC The present invention relates to human single exon nucleic acid probes  
 CC (SNP: see A110068-A1128459). The present sequence is a peptide encoded  
 CC by one such probe. The SENPs are derived from human Hela cells. The SENPs  
 CC can be used to produce a single exon microarray, which can be used for  
 CC measuring human gene expression in a sample derived from human cervical  
 CC epithelial cells. By measuring gene expression, the probes are therefore  
 CC useful in grading and/or staging of diseases of the cervix, notably  
 CC cervical cancer.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 110 AA;  
 Query Match 12.6%; Score 569; DB 22; Length 110;  
 Best Local Similarity 100.0%; Pred. No. 4.7e-34;  
 Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 568 OSDKFTAKPLDLSNGSRNDLNDRCGVPESASSEKAKEPTSDQTSSTESATNENNTNP 627  
 DB 1 OSDKFTAKPLDLSNGSRNDLNDRCGVPESASSEKAKEPTSDQTSSTESATNENNTNP 60  
 OY 628 EPQFQTEATGSAHEETSTRDSALQDTPDSDDDPVLIPGARYRAGPGR 676  
 DB 61 EPQFQTEATGSAHEETSTRDSALQDTPDSDDDPVLIPGARYRAGPGR 109  
 RESULT 14  
 AAM03874  
 ID AAM03874 standard; Protein; 110 AA.  
 AC AAM03874;  
 XX  
 DT 09-OCT-2001 (first entry)  
 DE Peptide #556 encoded by probe for measuring breast gene expression.  
 XX  
 KW Probe; human; breast disease; breast cancer; development disorder;  
 KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157270-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 29-JAN-2001; 2001WO-US00661.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 DR WPI; 2001-476286/51.  
 XX  
 PT Novel single exon nucleic acid probe used to measuring gene expression  
 PS in a human breast -  
 CC Claim 27; SEQ ID NO 12614; 322pp; English.  
 XX  
 CC The present invention relates to novel single exon nucleic acid probes  
 CC (see A110010-A1110067). The present sequence is a peptide encoded by one  
 CC such probe. The probes are useful for measuring human gene expression in  
 CC a human breast sample, where the probe hybridises at high stringency to a  
 CC nucleic acid expressed in the human breast. The probes are useful for  
 CC predicting, diagnosing, grading, staging, monitoring and prognosing  
 CC diseases of the human breast, particularly those diseases with polygenic  
 CC aetiology. The diseases include: breast cancer, disorders of development,  
 CC inflammatory diseases of the breast, fibrocystic changes, proliferative  
 CC breast disease and non-carcinoma tumours.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 110 AA;  
 Query Match 12.6%; Score 569; DB 22; Length 110;  
 Best Local Similarity 100.0%; Pred. No. 4.7e-34;  
 Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 568 OSDKFTAKPLDLSNGSRNDLNDRCGVPESASSEKAKEPTSDQTSSTESATNENNTNP 627  
 DB 1 OSDKFTAKPLDLSNGSRNDLNDRCGVPESASSEKAKEPTSDQTSSTESATNENNTNP 60  
 OY 628 EPQFQTEATGSAHEETSTRDSALQDTPDSDDDPVLIPGARYRAGPGR 676  
 DB 61 EPQFQTEATGSAHEETSTRDSALQDTPDSDDDPVLIPGARYRAGPGR 109  
 RESULT 15  
 ABG37880  
 ID ABG37880 standard; Peptide; 110 AA.  
 AC ABG37880;  
 XX  
 DT 19-AUG-2002 (first entry)  
 DE Human peptide encoded by genome-derived single exon probe SEQ ID 27545.  
 XX  
 KW Human; single exon probe; asthma; lung cancer; COPD; ILD;  
 KW chronic obstructive pulmonary disease; interstitial lung disease;  
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
 KW Hermansky-Rudrak syndrome; sarcoidosis; pulmonary haemosiderosis;  
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Kargener syndrome;  
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
 KW primary ciliary dyskinesia; pulmonary hypertension;  
 KW hyaline membrane disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200186003-A2.  
 XX  
 PD 15-NOV-2001.





Db 204 LHDQLENTAKCLAIINPRFTEYLAVGTNDPFAFVYDRRLPST-----NGNGISACVAYY 258  
 QY 245 IPSHLNKKSCR-----VTSICYSBDGQELIVSYSDIYLFDPKDDTARELKTPSAE 296  
 Db 259 AAGQIVKNISRNIVHEPRITVITFMNGTELLVNGICEHVYRFDLNH-----306  
 QY 297 ERREELROPVVKRLRLGDMWSDTGPRAPRESEBERDEQSPNVSLMQRSDMLSRWFEFA 356  
 Db 307 -----AEPV-----FYDLPAFSTLVHEE 326  
 QY 357 SEVAQNRGRGRSRPRGTSQSDISTLPVPSPDLEVESELAWEVTPAEOFLQPTSSST 416  
 Db 327 EPVKMPHRSR-----SLPT-----ELEVK-----KEGNDFFE-----354  
 QY 417 MSAQAHSTSPTEPHSTPLSSPDSEOROSVEASGHHTHOSDNNNEKLSPRKGTGEVY 476  
 Db 355 -----NGKLVDAIDAVSALAKYPOE-----Y 377  
 QY 477 LSLHSTEGTWTSTKLN-FTDEWSSIASSRGISGHCKSEGOESFVPOSSVQPEGDS 535  
 Db 378 LVLN-----RATALMRGWFGDYIALRDCHEAL-----RLDP-----410  
 QY 536 ERKAPESSEDTKYQEGVSAENPVENHINIQSDKFTAKPLDSNGERNDLNDSGCV 595  
 Db 411 -----SYVAHRFLA-----RALLELR-----428  
 QY 596 PESASSEKAKEPETSDQSTESATENNNTNPPQFOTEGATGSAHEETSTRDALSQDPTD 655  
 Db 429 PQDADCCLOALQORFDFANNHGVMLNDIKENRQSKS--PEAE-----LQPAY 478  
 QY 656 DSDDDPVLIPGARYRAGPDDRSAVARIOEFFRRKREKEMELDLNLRPLVKWYKYG 715  
 Db 479 VDD-----GFRYL-----RKKNAYDURSTARQDMQR-YVG 508  
 QY 716 HRNSRTIMEKANWGA--NFVMSGSDCGHLEFMDRTAEHLMLDEADNHVNCLOPHPD 773  
 Db 509 KCVNTTIDIEANLISGEGEIAAGSDGKNWYIEGDTGKRAYVRADSAIVNCVQHPST 568  
 QY 774 PILASSGIDYDIKIVSPLEESRIFNKKLADDEVITRNELMEETR 817  
 Db 569 CMLATSGIDHNKIKMSPCAAS-----AEE--RNLVADVTR 602  
 RESULT 17  
 AAU80462  
 ID AAU80462 standard; Protein; 628 AA.  
 AC AAU80462;  
 DT 12-MAR-2002 (first entry)  
 DE Fruit fly adipose protein, adp, #2.  
 KW Adipose protein; adp; obesity; transgenic animal; obesity;  
 KW adipositas; bulimia; wasting; cachexia; eating disorder;  
 KW body weight disorder; weight loss; cancer; infectious disease;  
 KW hypogonadism; Prader-Willi syndrome; Laurence-Moon-Biedl syndrome;  
 KW hypothyroidism; diabetes; Cushing's syndrome; endocrine disorder;  
 KW gastrointestinal diseases; inflammatory bowel disease;  
 KW ulcerative colitis; anorexia nervosa; glycogen storage disease;  
 KW lipid storage disease; lipoma; liposarcoma; heart disease; hypertension;  
 KW infertility; acquired immunodeficiency syndrome; AIDS.  
 XX Drosophila melanogaster.  
 OS  
 PN WO200196371-A2.  
 PD 20-DEC-2001.  
 PF 13-JUN-2001; 2001WO-EP06713.  
 PR 16-JUN-2000; 2000US-211914P.  
 PR 23-JUN-2000; 2000EP-0113049.

PR 28-JUN-2000; 2000US-214518P.  
 PR 17-APR-2001; 2001EP-0109537.  
 XX  
 PA (DEVE-) DEVELOGEN AG.  
 PI Broenner G, Ciossek T, Dohrmann C, Haeder T, Rothe M;  
 DR WPI; 2002-106464/14.  
 DR N-PSDB; ABR16345.  
 XX  
 PT Novel nucleic acid encoding adipose polypeptide which regulates, causes  
 PT or contributes to obesity, useful for treating obesity, heart disease,  
 PT hypertension, infertility, and controlling weight loss in cancer  
 PT patients -  
 PS  
 CC Claim 1; Page 142-144; 188pp; English.  
 CC  
 CC The invention relates to a nucleic acid encoding a adipose (ADP)  
 CC polypeptide which regulates, causes or contributes to obesity in an  
 CC animal or a human. The polynucleotides, proteins, anti-adp antibodies,  
 CC modulators of adp activity, adp antisense nucleic acids, expression  
 CC vectors, adp transgenic animals are useful in the diagnosis and  
 CC treatment of obesity, adipositas, bulimia, wasting (cachexia), eating  
 CC disorders and/or disorders of body weight/body mass, weight loss due to  
 CC cancer or infectious diseases, genetic disorders associated with  
 CC hypogonadism e.g. Prader-Willi syndrome, Laurence-Moon-Biedl syndrome,  
 CC hypothyroidism, diabetes, Cushing's syndrome, endocrine disorders,  
 CC gastrointestinal diseases, inflammatory bowel disease, ulcerative  
 CC colitis, and anorexia nervosa. They are also useful for treating  
 CC disorders of body weight/mass e.g. glycogen storage diseases, and lipid  
 CC storage diseases and for treating lipomas, and/or liposarcomas. The  
 CC compositions are also useful for treating heart disease, hypertension,  
 CC and infertility and for treating conditions associated with under weight  
 CC e.g. enhancing or controlling fertility, controlling weight loss in  
 CC acquired immunodeficiency syndrome (AIDS) or cancer patients. The  
 CC present sequence is an adp protein.  
 XX  
 SQ Sequence 628 AA;  
 Query Match 11.2%; Score 508.5; DB 23; Length 628;  
 Best local Similarity 23.2%; Pred. No. 1.8e-28;  
 Matches 191; Conservative 107; Mismatches 261; Indels 265; Gaps 26;  
 QY 10 LMDVRRKSLGLDPSRLNSRYLGRREFIQRLKLEATLVNHDGCVNTICWDTGEXYLSG 69  
 Db 28 LHM--ORQGHLELDEMLRRRLASPAYVDRLBOEAVLVGHECVCLLEWTDGMMLASG 85  
 QY 70 SDDTKLVISNPYRKVLITIRSGHRANIPSAKFLPCTNDKQIYVSGSDGYIFFTNWEGDA 129  
 Db 86 SDDYRWIMDPFRKKLVHIVIRTHLGVESVFKFLPTNNSIVATCAADKFIYYVDINDPN 145  
 QY 130 ETRNOCFTCHYGTTEIMTVPNDPYTFLSGEGGTWRWMDTRIKTSCREDKDDILIN 189  
 Db 146 ETLPSC--ICHFSRAKRLAIAQDSPIVFWAGSDGCLQDIDIEPRRCREBIEIGVRLN 203  
 QY 190 C-----RRAATVSLACPPIDPYLVAVGSSSVRIYDRMLGTRATGNYAGRTTGVARF 244  
 Db 204 LHDQLENTAKCLAIINPRFTEYLAVGTNDPFAFVYDRRLPST-----NGNGISACVAYY 258  
 QY 245 IPSHLNKKSCR-----VTSICYSBDGQELIVSYSDIYLFDPKDDTARELKTPSAE 296  
 Db 259 AAGQIVKNISRNIVHEPRITVITFMNGTELLVNGICEHVYRFDLNH-----306  
 QY 297 ERREELROPVVKRLRLGDMWSDTGPRAPRESEBERDEQSPNVSLMQRSDMLSRWFEFA 356  
 Db 307 -----AEPV-----FYDLPAFSTLVHEE 326  
 QY 357 SEVAQNRGRGRSRPRGTSQSDISTLPVPSPDLEVESELAWEVTPAEOFLQPTSSST 416  
 Db 327 EPVKMPHRSR-----SLPT-----ELEVK-----KEGNDFFE-----354  
 QY 417 MSAQAHSTSPTEPHSTPLSSPDSEOROSVEASGHHTHOSDNNNEKLSPRKGTGEVY 476

Db 355 -----NGKLVDALDAYSAALAKYPOGE-----Y 377

Qy 477 LSLHSTBETTTSTIKLN-FTEWMSIASSRGIGSHCKSEGOEESFVQSVQPEGDS 535

Db 378 LYLN-----RATALLRRGFGDIYALRDCHEAL-----RLDP-----410

Qy 536 ETKAPEESEDVTKYQEGVSAENPVENHINITQSDFTAKPLDLSNGERNDLMDSCGV 595

Db 411 -----SYKAHFRLLA-----RALLELR-----428

Qy 596 PESASSEKAKPEPSTSDQSTESATENNENTNEPQFOTATGSAHESTSDSALQDTD 655

Db 429 PQDAPQCCALQALQRFPPDFANNHGVLMNDIKENRQSKS--PEAE-----LQPE 478

Qy 656 DSDDDPVILPCARYAGPGRDSARVARIOEFRRRERKEMELDTLNRPLVKMYKG 715

Db 479 VDD-----GFRYL-----RLKMEYDTRSTARDWQR--YVG 508

Qy 716 HRNSTMIKEANFWGA--NFVMSGSDCHIFIMDRHTAHMLLEADNHVYVNCLOPPFD 773

Db 509 HCNVTTDIKEANYLSSGGEFTIAGSDGNNMTWEGDTGKIRAVYRADSAIVNCVQPHPI 568

Qy 774 PILASGIDYDIKIMSPLEESRIFNRKLADEVITRNEMLLETR 817

Db 569 CMLATSGIDHNKIMSPCAS-----AE-----RNLVADYTR 602

RESULT 18

AAU80464

ID AAU80464 standard; Protein; 677 AA.

AC AAU80464;

DT 12-MAR-2002 (first entry)

DE Human adipose protein, adp.

XX

KW Adipose protein; adp; obesity; transgenic animal; obesity; adipositas; bulimia; wasting; cachexia; eating disorder; body weight disorder; weight loss; cancer; infectious disease; hypogonadism; Prader-Willi syndrome; Laurence-Moon-Biedl syndrome; hypothyroidism; diabetes; Cushing's syndrome; endocrine disorder; gastrointestinal diseases; inflammatory bowel disease; ulcerative colitis; anorexia nervosa; glycogen storage disease; lipid storage disease; lipoma; liposarcoma; heart disease; hypertension; infertility; acquired immunodeficiency syndrome; AIDS.

OS Homo sapiens.

PN WO200196371-A2.

PD 20-DEC-2001.

PF 13-JUN-2001; 2001WO-EP06713.

PR 16-JUN-2000; 2000US-211914P.

PR 23-JUN-2000; 2000EP-0113049.

PR 28-JUN-2000; 2000US-214518P.

PR 17-APR-2001; 2001EP-0109537.

PA (DEVE-) DEVELOPEN AG.

PI Broenner G, Ciossek T, Dohrmann C, Haeder T, Nothe M;

DR WPI; 2002-106464/14.

DR N-PSDB; ABK16347.

XX

PT Novel nucleic acid encoding adipose polypeptide which regulates, causes or contributes to obesity, useful for treating obesity, heart disease, hypertension, infertility, and controlling weight loss in cancer patients -

PT Claim 1; Page 153-155; 188pp; English.

XX The invention relates to a nucleic acid encoding a adipose (ADP) polypeptide which regulates, causes or contributes to obesity in an animal or a human. The polynucleotides, proteins, anti-adp antibodies, CC modulators of adp activity, adp antisense nucleic acids, expression CC vectors, adp transgenic animals are useful in the diagnosis and CC treatment of obesity, adipositas, bulimia, wasting (cachexia), eating CC disorders and/or disorders of body weight/body mass, weight loss due to CC cancer or infectious diseases, genetic disorders associated with CC hypogonadism e.g. Prader-Willi syndrome, Laurence-Moon-Biedl syndrome, CC hypothyroidism, diabetes, Cushing's syndrome, endocrine disorders, CC gastrointestinal diseases, inflammatory bowel disease, ulcerative CC colitis, and anorexia nervosa. They are also useful for treating CC disorders of body weight/mass e.g. glycogen storage diseases, and lipid CC storage diseases and for treating lipomas, and/or liposarcomas. The CC compositions are also useful for treating heart disease, hypertension, CC and infertility and for treating conditions associated with under weight CC e.g. enhancing or controlling fertility, controlling weight loss in CC acquired immunodeficiency syndrome (AIDS) or cancer patients. The CC present sequence is an adp protein.

XX

Sequence 677 AA;

Query Match 11.0%; Score 499.5; DB 23; Length 677;  
Best Local Similarity 23.6%; Pred. No. 9.4e-28;  
Matches 209; Conservative 116; Mismatches 297; Indels 265; Gaps 33;

Qy 13 DVRRSLGLEDPRLSRSTYLGREFIORKLEATLVNHDGCVNTICMNDTGEYISGSD 72

Db 9 DLIRQIKERALSFRERYHTDPTIRRLGELAELOHGSCVNCLEWNEKGDILASGSD 68

Qy 73 TKLVISNPSYKRVLTITRSGHRANIFSAFLPCTNDKQIVSCSGGVIFYTEVEDAEIN 132

Db 69 QHITWDEPLHKKILSMHTGTANIFSVKFLPHAGDRILITGAADSKVHVDLYETIH 128

Qy 133 RQCFCHYCTTEIMYVNDPYTFLSCGSDGVWMPDTRITSCTECKC--DQILIN-- 189

Db 129 --MGDTNRVKKRATAPMMPNFWMSAEDLIDQYDLR-----ENSKASEVILDTL 178

Qy 190 --CRR--AATSVATCPPIPYLVAGCSDSVRIYDRML-----GTRATGNYVRGT-- 237

Db 179 EYCGQIVAKCLTVNPQNNCLAVASGPFVYLYIRIMHNHRKMSKQSPSAGVHTFCR 238

Qy 238 -----TGMAVAFIPSHL-----NNKSCRV--TSLCYSEDOEILVSYSDIYLFDPK 283

Db 239 QKPLPDGAQYVAGHLPVKLPDYNRRLVAVATVTFSPNGETELLVNNGGQVYLF-- 296

Qy 284 DDTARELKTPEAEERRELRQPPVYKRLRLRGWSDTPRARERERERGBSPNVSLMQ 343

Db 297 -----LTYKORPYFTL-----PRKHSSEGVQNGKMTN----- 326

Qy 344 RMSDLSRWFEEASEVAVASNGRGRSRPRGTSQSDISTLPVVPSPDLE--VSETAMEVD 402

Db 327 GVSNGVSGNLHNSGFLPESRGHVSPQ-----VELPPLYERKQOQANEA-- 372

Qy 403 TPAQOFLQPTSSYMSAQAHSSTSPESHPTPLSSPDSEORQSVASGHTTHQSDNN 462

Db 373 FACQW-----TQALDLYS--KAVQARAHNMALYG-----NR 402

Qy 463 NEXLSRPGGEPVLSLHYSTEGTTSITIKLNFTEWMSIASSRGIGSHCKSEGOEESF 522

Db 403 AAAYMKRMQSD-----HYDALRDLCAKATSLN-----PCHLKA----- 435

Qy 523 VPOSSVQPEEGSETKAPEESEDVTKYQEGVSAENPVENHINITQ--SDFTAKPLDS 579

Db 436 -----HRLARCLPELKYVAEAL-- 453

Qy 580 NSGERNDLNDKSCGVPEASASEKAKPEPSTSDQSTESATENNENTNEPQFOTATGSPS 639

Db 454 -----ECIDDFKGRPEQAHSSA--CDALG-- 476

Qy 640 AHETSTRD--SALQDIDDSDDDPVLLPGAVYRAGPGRSARVARIOEFRRRKRKKEME 697

Db 477 -----RDITAAFSKNDGE-----KKGFG--GAPVRL-----RSTSRKDSI 512  
 QY 698 ELDTINIRPL--VKAVYKGRHSRTMIKEANFWGAN--FVMSGSDGHIIFMDRTAAH 753  
 Db 513 SDEWVLRERSYDQFRYCGHCTTTDKEANFPFGSAGIYVSGSDGSFFIWEKETTLM 572  
 QY 754 LMLEADNHVNCLOPHPPDPIILASSGIDYDIKIMSPLEESRIENKRLADEVITRNEML 813  
 Db 573 VAVLGDESIVNCLOPHPSYCFIATSGIDPVVRLMNPRESSEDLTGAVVEDMEGASQ--- 629  
 QY 814 EETRTITVPAFMLRLMLNHIRADRLGDSRSGSGGGENENED 860  
 Db 630 ---ANORRNADPLEYMLNMGi---KITGLSSGAGASDDSDSE 669  
 RESULT 19  
 AAU80463  
 ID AAU80463 standard; Protein; 677 AA.  
 AC AAU80463;  
 DT 12-MAR-2002 (first entry)  
 DE Mouse adipose protein, adp.  
 XX  
 XX Adipose protein; adp; obesity; transgenic animal; obesity;  
 KM adipositas; bulimia; wasting; cachexia; eating disorder;  
 KM body weight disorder; weight loss; cancer; infectious disease;  
 KM hypogonadism; Prader-Willi syndrome; Laurence-Moon-Biedl syndrome;  
 KM hypothyroidism; diabetes; Cushing's syndrome; endocrine disorder;  
 KM gastrointestinal diseases; inflammatory bowel disease;  
 KM ulcerative colitis; anorexia nervosa; glycogen storage disease;  
 KM lipid storage disease; lipoma; liposarcoma; heart disease; hypertension;  
 KM lipid storage; acquired immunodeficiency syndrome; AIDS.  
 XX  
 OS Mus musculus.  
 XX  
 XX WO200196371-A2.  
 XX  
 XX PD 20-DEC-2001.  
 XX  
 XX 13-JUN-2001; 2001WO-EP06713.  
 XX  
 XX 16-JUN-2000; 2000US-211914P.  
 XX PR 23-JUN-2000; 2000EP-0113049.  
 XX PR 28-JUN-2000; 2000US-214518P.  
 XX PR 17-APR-2001; 2001EP-0109537.  
 XX  
 XX (DEVE-) DEVELOGEN AG.  
 XX  
 XX Broenner G, Ciossek T, Dohrmann C, Haeder T, Rothe M;  
 PI WPI: 2002-106464/14.  
 XX  
 XX DR N-PSDB; ABR16346.  
 XX  
 XX Novel nucleic acid encoding adipose polypeptide which regulates, causes  
 PT or contributes to obesity, useful for treating obesity, heart disease,  
 PR hypertension, infertility, and controlling weight loss in cancer  
 PI patients -  
 XX  
 XX Claim 1; Page 148-150; 188pp; English.  
 XX  
 XX The invention relates to a nucleic acid encoding a adipose (ADP)  
 CC polypeptide which regulates, causes or contributes to obesity in an  
 CC animal or a human. The polynucleotides, proteins, anti-adp antibodies,  
 CC modulators of adp activity, adp antisense nucleic acids, expression  
 CC vectors, adp transgenic animals are useful in the diagnosis and  
 CC treatment of obesity, adipositas, bulimia, wasting (cachexia), eating  
 CC disorders and/or disorders of body weight/body mass, weight loss due to  
 CC cancer or infectious diseases, genetic disorders associated with  
 CC hypogonadism e.g. Prader-Willi syndrome, Laurence-Moon-Biedl syndrome,  
 CC hypothyroidism, diabetes, Cushing's syndrome, endocrine disorders,  
 CC gastrointestinal diseases, inflammatory bowel disease, ulcerative

CC colitis, and anorexia nervosa. They are also useful for treating  
 CC disorders of body weight/mass e.g. glycogen storage diseases, and lipid  
 CC storage diseases and for treating lipomas, and/or liposarcomas. The  
 CC compositions are also useful for treating heart disease, hypertension,  
 CC and infertility and for treating conditions associated with under weight  
 CC e.g. enhancing or controlling fertility, controlling weight loss in  
 CC acquired immunodeficiency syndrome (AIDS) or cancer patients. The  
 CC present sequence is an adp protein.  
 XX  
 SQ Sequence 677 AA:  
 Query Match 10.8%; Score 489.5; DB 23; Length 677;  
 Best Local Similarity 23.4%; Pred. No. 5.1e-27;  
 Matches 208; Conservative 113; Mismatches 298; Indels 271; Gaps 32;  
 QY 13 DYKRSLGLEDEPSRLSRSLGRREFIQRKLEATLVNHDGCVNTICMNDTGEYILSGSD 72  
 Db 9 DLIRKQYKERGALSFERRHYVDPFIRRLGLEAELOGHSGCVNCLMNERGDLILASGSD 68  
 QY 73 TKLIVSNPYSRKVLTTIRSGHRANIPSAFLPCTNDKQIVSGSGDVIFEYTVNEQDAETN 132  
 Db 69 QHTIYVMDPLHKKLLSMHTGHTANIFSVKFLPHAGDRILITGAADSKVYHDLTVAEIHH 128  
 QY 133 ROCQFCHYGTTEIYNTVNDPTPLSCGEDGVKRFDRIKTSCKEDCK-DIILINCR 191  
 Db 129 ---MFGDHTNRVKRIATAPMWPTEFWSAEDGILRQYDR-----ENSKSEVILIDL 178  
 QY 192 R-----AATSAICPIPIYUAVGSDSVYRILYDRML-----GTRATGNVAGRG----- 237  
 Db 179 ETCGPVVEAKCLTVNQDNNAICAVGASGPFVRLIDRIIMHNRKSKROSPACVHTFCOR 238  
 QY 238 ----TGVAARFIPSHL-----NNKSCNV--TSLCYSEDOGEILVSYSDYILFDPK 283  
 Db 239 QKPLPGAAQYVAGHLPKVLPDYNRLRYLVATYVTFSPNGTELLVNMKGEGVYLFED-- 296  
 QY 284 DDTARELKTPSAREERREELROPVYKRLRLGMSDIPRARPSEREERGEOESPNNVSLMQ 343  
 Db 297 -----LTYKQRPYFLL-----PR-KCHSVQVQNGKSTN----- 325  
 QY 344 RMSDMLRWFEEASEVAQSGNRGRSRPRGTSQSDISTLPYSPSPDLE-VSETIMEYD 402  
 Db 326 GVSNGVSNGLHLHS-----NGFRLPK-----SKGIS--POVELPYLERVQAQNEA- 371  
 QY 403 TPAPGFLDPSTSTMSQAQHSSTSPSPSPHSPPLSSPDEQRQSVASGSHHTHOSDNN 462  
 Db 372 FACQGM-----TQAIQLYS-QAVQKAPHNALYg-----NR 401  
 QY 463 NEKLSPKPGTGEVYLSLHSTEGTTSTIKINTDMSSIASSRGIGSHCKSEGOBESR 522  
 Db 402 AAAYMKRKWDG-----HYDALRDLKAISLN-----PCHLKA----- 434  
 QY 523 VPOSSVQPPGDESETKAPESSEEDVTKYQEGVSAENPVENHINITQ---SDKFTAKPLDS 579  
 Db 435 -----HFLACLCELEKVAAL-- 452  
 QY 580 NGERNDLINDRSGVPEESASSEKAKEPETSQDSTE-----SATNENNTNPPQFO 632  
 Db 453 -----ECIDDFKGFPEQAHSSACDALGRDITAAFLRSKSGEEKKAA 494  
 QY 633 TEATGSAHEHSTRDLSAQDTDDSDDPVLPFGATYRAGPGRRSAVARIOEFFRRKE 652  
 Db 495 GGGGGFVRLRSTSRKDSI-----SEDEVL-----RERSV 524  
 QY 693 RKEMEELDTLIRPLVKKMYKGRHSRTMIKEANFWGAN--FVMSGSDGHIIFMDRTAAH 750  
 Db 525 DYQFR-----YCHCHCTTTDKEANFPFGSAGIYVSGSDGSFFIWEKETT 569  
 QY 751 AEHLMLLEADNHVNCLOPHPPDPIILASSGIDYDIKIMSPLEESRIENKRLADEVITRNE 810  
 Db 570 TNLVAVLGDESIVNCLOPHPSYCFIATSGIDPVVRLMNPRESSEDLTGAVVEDMEGASQ 629  
 QY 811 LMLEETRNITVPAFMLRLMLNHIRADRLGDSRSGSGGGENENED 860





PA (DEVE-) DEVELOGEN AG.  
 XX  
 PI Broenner G, Closssek T, Dohmann C, Haeder T, Rothe M;  
 XX WPI; 2002-106464/14.  
 DR  
 XX Novel nucleic acid encoding adipose polypeptide which regulates, causes  
 PT or contributes to obesity, useful for treating obesity, heart disease,  
 PT hypertension, infertility, and controlling weight loss in cancer  
 PT patients -  
 XX  
 PS Disclosure; Page 165-167; 188pp; English.  
 XX  
 CC The invention relates to a nucleic acid encoding a adipose (ADP)  
 CC polypeptide which regulates, causes or contributes to obesity in an  
 CC animal or a human. The polynucleotides, proteins, anti-adp antibodies,  
 CC modulators of adp activity, adp antisense nucleic acids, expression  
 CC vectors, adp transgenic animals are useful in the diagnosis and  
 CC treatment of obesity, adipositas, bulimia, wasting (cachexia), eating  
 CC disorders and/or disorders of body weight/body mass, weight loss due to  
 CC cancer or infectious diseases, genetic disorders associated with  
 CC hypogonadism e.g. Prader-Willi syndrome, Laurence-Moon-Biedl syndrome,  
 CC hypothyroidism, diabetes, Cushing's syndrome, endocrine disorders,  
 CC gastrointestinal diseases, inflammatory bowel disease, ulcerative  
 CC colitis, and anorexia nervosa. They are also useful for treating  
 CC disorders of body weight/mass e.g. glycogen storage diseases, and lipid  
 CC storage diseases and for treating lipomas, and/or liposarcomas. The  
 CC compositions are also useful for treating heart disease, hypertension,  
 CC and infertility and for treating conditions associated with under weight  
 CC e.g. enhancing or controlling fertility, controlling weight loss in  
 CC acquired immunodeficiency syndrome (AIDS) or cancer patients. The  
 CC present sequence is an adp associated protein included in the  
 CC sequence listing but not mentioned any where else in the specification.  
 CC  
 XX  
 SQ Sequence 580 AA:  
 Query Match 10.2%; Score 460.5; DB 23; Length 580;  
 Best Local Similarity 24.1%; Pred. No. 5.4e-25;  
 Matches 192; Conservative 92; Mismatches 252; Indels 261; Gaps 30;  
 QY 40 RLKLEATLVHDCGVNTICWMDTGEYILSGSDPTFKLVISNPSKRVLTIRSGHRANIPS 99  
 DB 1 RLGLFALOGHSGCVNCLNEMKEKDLASGSDDTIYWDPLHKKLSMHTGHRANIPS 60  
 QY 100 AKFLPCTNDKQIVSCSGDGVIFTYNVEDAETNRQCOFTCHYGTYTEIMTVNDPYTFLS 159  
 DB 61 VKFLPHAGDRILITGADSKVHVDLVYKETIH--MGDHTNRYKRLATAMWNTFMS 117  
 QY 160 CGEDGYRWFDTRIKTSCTKEDCK-DLLINCR-----AATSAICPPIRYIAVGCS 212  
 DB 118 AAEGLLROYDLR-----ENSKHSEVLIDLTEXCGPMEVAKCLTVNDNNCLAVGAS 170  
 QY 213 DSSVRIYDRML-----GTRATGNYAGRG-----TGMVARIPLSHL-----NNKS 253  
 DB 171 GPVRLIDINHHKRSKMSQSPAGVHTPCDRQKPLPDGAQYVAGHLPKLPDYNSR 230  
 QY 254 CRV---TSLCYSEDEGILVYSDDIYLFDPKDTARELKTPSAERREELROPPVKRL 310  
 DB 231 LRVLVATYTFESPNGTELVVMGEOYVYLF-----LTYKQRYPTFL 272  
 QY 311 RLKGDMSDTGPRARPESEEREDGESPVSILMQRASDMLSRFEEASEVAQSNRGGRSR 370  
 DB 273 L-----PR-KCHSEVVGNGKMNSTN-----GVSNGVSNGLHLS-----NGFRL 309  
 QY 371 PRGTSOSDISTLPTVSSPPLP-EVSETAMEVDPFAEOLFDPSTSTSSAQAHSSTSPPE 429  
 DB 310 PE---SKGCS--PQVELPYLERKQOANA-PAQOQ-----TQITQLYS--QAVOK 355  
 QY 430 SPHSTPLSPDSERQSVASGHTHQSNDNNNEKLSPKPGTGPVLSLHSTEGTSTS 489  
 DB 356 APHNAMLXG-----NRAAAVMKRMKWDG-----HYDALRDCIK 388  
 QY 490 TIKLFTDEMSSIASSSRGISGCKSGEGESFVQSSVQPPEDGETKAPRESSSEDTVK 549

DB 389 AISLN-----PCHLKA----- 399  
 QY 550 YQGVSAENFVENHINITO---SDKFTAKPLDSSNGERNDLNDRCGVPPESSASSEKAK 606  
 DB 400 -----HFRRLARCLFELKYVAEL-----ECLDLDFGK 426  
 QY 607 EPETSDQTSNE-----SATINENNTNPEPOFQTEATGPSAHEETSTRDSALQDTDSDD 659  
 DB 427 FPEQAHSSACDALGRDITTAIFSRSKDEEKKAAAGGGGPPVRLRSTRKDSI-----SED 480  
 QY 660 DPAVLICARRRAGRGDRSARVARIQEFFRRRKEKKEHEEDLTNIRRLVAMVYKGRHNS 719  
 DB 481 EMVL-----RERSYDQFR-----YCGHONT 501  
 QY 720 RTMIKEANFNGAN--FVMSGSDCGHIFIPDRHTAEHMLLEADNHVYVNCLOPHFPDPIIA 777  
 DB 502 TTQIKENFNGSNAGQIVYSGSDOSFFIWEKETINILRVVLQGDGSIYVNCLOPHRSYCLA 561  
 QY 778 SSGIDVDIKTWSPLEES 794  
 DB 562 TSGIDPVRRLMNPRES 578  
 RESULT 23  
 AA080471  
 ID AAU80471 standard; Protein; 616 AA.  
 XX  
 AC AAU80471;  
 XX  
 DT 12-MAR-2002 (first entry)  
 XX  
 DE Mouse adipose protein, adp. #2.  
 XX  
 KW adipose protein; adp; obesity; transgenic animal; obesity;  
 KW adipositas; bulimia; wasting; cachexia; eating disorder;  
 KW body weight disorder; weight loss; cancer; infectious disease;  
 KW hypogonadism; Prader-Willi syndrome; Laurence-Moon-Biedl syndrome;  
 KW hypothyroidism; diabetes; Cushing's syndrome; endocrine disorder;  
 KW gastrointestinal diseases; inflammatory bowel disease;  
 KW ulcerative colitis; anorexia nervosa; glycogen storage disease;  
 KW lipid storage disease; lipoma; liposarcoma; heart disease; hypertension;  
 KW infertility; acquired immunodeficiency syndrome; AIDS.  
 XX  
 OS Mus musculus.  
 OS  
 PN W0200196371-A2.  
 XX  
 PD 20-DEC-2001.  
 XX  
 PF 13-JUN-2001; 2001WO-EP06713.  
 XX  
 PR 16-JUN-2000; 2000US-211914P.  
 PR 23-JUN-2000; 2000EP-0113049.  
 PR 28-JUN-2000; 2000US-214518P.  
 PR 17-APR-2001; 2001EP-0109537.  
 XX  
 PA (DEVE-) DEVELOGEN AG.  
 XX  
 PI Broenner G, Closssek T, Dohmann C, Haeder T, Rothe M;  
 DR WPI; 2002-106464/14.  
 DR N-PSDB; ABK16384.  
 XX  
 PT Novel nucleic acid encoding adipose polypeptide which regulates, causes  
 PT or contributes to obesity, useful for treating obesity, heart disease,  
 PT hypertension, infertility, and controlling weight loss in cancer  
 PT patients -  
 PS Claim 1; Page 175-177; 188pp; English.  
 CC The invention relates to a nucleic acid encoding a adipose (ADP)  
 CC polypeptide which regulates, causes or contributes to obesity in an

CC animal or a human. The polynucleotides, proteins, anti-adp antibodies,  
 CC modulators of adp activity, adp antisense nucleic acids, expression  
 CC vectors, adp transgenic animals are useful in the diagnosis and  
 CC treatment of obesity, adipositas, bulimia, wasting (cachexia), eating  
 CC disorders and/or disorders of body weight/body mass, weight loss due to  
 CC cancer or infectious diseases, genetic disorders associated with  
 CC hypogonadism e.g. Prader-Willi syndrome, Laurence-Moon-Biedl syndrome,  
 CC hypothyroidism, diabetes, Cushing's syndrome, endocrine disorders,  
 CC gastrointestinal diseases, inflammatory bowel disease, ulcerative  
 CC colitis, and anorexia nervosa. They are also useful for treating  
 CC disorders of body weight/mass e.g. glycogen storage diseases, and lipid  
 CC storage diseases and for treating lipomas, and/or liposarcomas. The  
 CC compositions are also useful for treating heart disease, hypertension,  
 CC and infertility and for treating conditions associated with under weight  
 CC e.g. enhancing or controlling fertility, controlling weight loss in  
 CC acquired immunodeficiency syndrome (Aids) or cancer patients. The  
 CC present sequence is an adp protein.

XX Sequence 616 AA;

Query Match 9.7%; Score 441; DB 23; Length 616;

Best Local Similarity 21.9%; Pred. No. 1.6e-23;

Matches 191; Conservative 111; Mismatches 275; Indels 294; Gaps 30;

QY 13 DYKRSGLGLEDPSRLSRSLRGLRRETIQRLKLEATLVHDGCVNTICWDTGETIISGSD 72  
 DB 9 DLIRROVKERKALSFERRHYVTDPEIRRLGLEAELOGSHGCVNCEMNEKGDLLASGSD 68  
 QY 73 TKLVSNPYSRKVLVTIRSGHRANFSAKFLPCTNDKOIVSGSGGVIFVTVEDEDAEIN 132  
 DB 69 QHTIWDPLHKKLSMHTGHANFSAKFLPCTNDKOIVSGSGGVIFVTVEDEDAEIN 128  
 QY 133 ROCQFCHYGTYTEIMTVPNDFELSCGDEGVNMF-----DTRIKTSCTKEDCKDIL 187  
 DB 129 ---MGGDHNRVRIATAPMNTFWSAEDGLIRKSMQSPAGVHTCDQ----- 178  
 QY 188 INCRRAASVAICPPPIYTLAVGCSDSVRYTRDRMLGTRANGTAGKGTGMVAFIDS 247  
 DB 179 -----KPLP-----DGAQYV-----VAGHLPV 196  
 QY 248 HLNKSCRY-----TSLCYSEDOELIVSSDYTYLPDPKDOTARELTPSAERREEL 302  
 DB 197 KLPTDINSRLRVATVYTFSPNGTELVNMGEOYTLFP-----LTY 238  
 QY 303 KOPVYKRLRLNGDWDGTPRARERERERDQSPNVSLMOMSDMLSNWFEASEVAOS 362  
 DB 239 KOPYTFLL-----PR-KCHSEVQNGKMTN-----GVNKGVSGLHLHS----- 278  
 QY 363 NRGRRSRPRGTSOSDSTLTPTVSSPDL-VSETAMEVDTAPAOFLQPTSSITMSAOA 421  
 DB 279 ---NGFRLEP-----SKGCS--POVELPPLYERVKOAOANA-PACQW-----TQAIOL 322  
 QY 422 HSTSPTESPHSTPLISSPDSBOQSVASGHTTHQSDNNNEKLSPKDGEPLVSLHY 481  
 DB 323 YS-QAVQKAPHNMLYG-----NRAAYTKRKWDG-----HY 354  
 QY 482 STEGTTSTIKLFTDEWSSIASSSRGIGSHCKSGEQESFVPOSSVOPPEGDSKAP 541  
 DB 355 DALRDLKALISLN-----PCHLKA----- 373  
 QY 542 ESSEVYTKQEGVSAENVEENHINITQ---SPKFTAKPLDSNGSERNDLNDRCGVPRE 598  
 DB 374 -----HFRILARCLFELKLYAEAL-----E 392  
 QY 599 SASSEKAKEPETSDQSTTE-----SATNENNTNPEQFOTEAATGPAHSEFTSRDASL 651  
 DB 333 CLDDPKGFPPEGAHSSACDALGRDITALLFSKSDGEKKAAGGGGPPVRLRSTSRDST- 451  
 QY 652 QDTDDSDDDPVLPAGARYRAGPGRRSAAVARIOEFFRRKKEKEMEDLTTLIRPLVM 711  
 DB 452 -----SEDEMYL-----RESYVYQYR----- 468  
 QY 712 YKGRHSRRTMKEANFNGAN--FWMSGSDCGHIFTWRHTAEHLMLLEADNHNVCLOP 769

DB 469 -CGHCNTTIDIKAEANFESNAQYIVSGSDGSEFFIWEKETTULVVLGDESVNCLP 527  
 QY 770 HEPDPLASSGIDYDIKISPLEESRIRNRKLADEVITNEMLETRTITVYPAFMLR 829  
 DB 528 HSYCYLATSIGDIPVRLNPNPRESEDLTGRAVEDMEGASQ-----ANORRNANPLEA 581  
 QY 830 MLASLNHIRADRLGDRSGSGOENENDEE 860  
 DB 582 MLMDKGY-----RITGLSSGAGASDDEDSAE 608

# RESULT 24

ABB27840

ID ABB27840 standard; Peptide; 86 AA.

AC ABB27840;

DT 01-FEB-2002 (first entry)

DE Human peptide #491 encoded by breast cell single exon nucleic acid probe.

XX Human; microarray; single exon probe; gene expression; breast;

KW disease; cancer.

XX Homo sapiens.

PN WO200157271-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00662.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

PT WPI; 2001-496933/54.

PS New spatially-addressable set of single exon nucleic acid probes,

XX useful for measuring gene expression in sample derived from human

XX breast, comprises number of single exon nucleic acid probes

XX Claim 27; SEQ ID NO 10808; 327bp + sequence listing; English.

XX The invention relates to a spatially-addressable set of single exon

XX nucleic acid probes for measuring gene expression in a sample derived

XX from human breast and BT 474 cells. The method involves contacting

XX the probes with a collection of detectably labeled nucleic acids

XX derived from mRNA of human breast, and then measuring the label

XX bound to each probe of the microarray. The probes are useful for

XX verifying the expression of regions of genomic DNA predicted to

XX encode proteins. They are useful for gene discovery, and for

XX determining predisposition and/or prognosing breast disease. Gene

XX expression analysis is useful for assessing the toxicity of chemical

XX agents on cells. The microarray of this invention presents a far greater

XX diversity of probes for measuring gene expression, with far less bias

XX than expressed sequence tag microarrays. The method is suitable for

XX rapid production of functional information from genomic sequence. The

XX present sequence is a peptide encoded by a single exon nucleic acid

XX probe of the invention.

XX Note: The sequence data for this patent did not form part of the

XX printed specification, but was obtained in electronic format directly

XX from Wipo at ftp.wipo.int/pub/published\_pcl\_sequences.



SQ Sequence 86 AA;

Query Match 9.7%; Score 437; DB 22; Length 86;  
Best Local Similarity 100.0%; Pred. No. 1.5e-24;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 374 GTSQSDISTLPVPSPPDLEVESETAMEVDTPAEQFLQPSSTMSQAQAHSTSSPTESPHS 433  
DB 1 GTSQSDISTLPVPSPPDLEVESETAMEVDTPAEQFLQPSSTMSQAQAHSTSSPTESPHS 60

OY 434 TPLISSPDSERQSVASGHHTHQS 459  
DB 61 TPLISSPDSERQSVASGHHTHQS 86

RESULT 25  
ABB33011  
ID ABB33011 standard; Peptide; 86 AA.  
XX  
AC ABB33011;  
XX  
DT 04-FEB-2002 (first entry)  
XX  
DE Peptide #517 encoded by human foetal liver single exon probe.  
XX  
KW Human; foetal liver; gene expression; single exon nucleic acid probe.  
XX  
OS Homo sapiens.  
XX  
PN WO200157277-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00669.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
PS WPI; 2001-483447/52.  
XX  
DR WPI; 2001-483447/52.  
XX  
PT Human genome-derived single exon nucleic acid probes useful for  
PS analyzing gene expression in human foetal liver -  
XX  
PS Claim 27; SEQ ID NO 25646; 639bp + sequence listing; English.  
XX  
CC The invention relates to a single exon nucleic acid probe for  
CC measuring human gene expression in a sample derived from human foetal  
CC liver. The single exon nucleic acid probes may be used for predicting,  
CC measuring and displaying gene expression in samples derived from human  
CC foetal liver. The present sequence is a peptide encoded by a single exon  
CC nucleic acid probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 86 AA;

Query Match 9.7%; Score 437; DB 22; Length 86;  
Best Local Similarity 100.0%; Pred. No. 1.5e-24;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 374 GTSQSDISTLPVPSPPDLEVESETAMEVDTPAEQFLQPSSTMSQAQAHSTSSPTESPHS 433  
DB 1 GTSQSDISTLPVPSPPDLEVESETAMEVDTPAEQFLQPSSTMSQAQAHSTSSPTESPHS 60

OY 434 TPLISSPDSERQSVASGHHTHQS 459  
DB 61 TPLISSPDSERQSVASGHHTHQS 86

RESULT 26  
ABB18484  
ID ABB18484 standard; Protein; 86 AA.  
XX  
AC ABB18484;  
XX  
DT 23-JAN-2002 (first entry)  
XX  
DE Protein #483 encoded by probe for measuring heart cell gene expression.  
XX  
KW Human; gene expression; heart; microarray; vascular system;  
KW cardiovascular disease; hypertension; cardiac arrhythmia;  
KW congenital heart disease.  
XX  
OS Homo sapiens.  
XX  
PN WO200157274-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00666.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
PS WPI; 2001-488899/53.  
XX  
DR WPI; 2001-488899/53.  
XX  
PT Single exon nucleic acid probes for analyzing gene expression in human  
PT hearts -  
XX  
PS Claim 15; SEQ ID NO 20254; 530bp; English.  
XX  
CC The present invention relates to single exon nucleic acid probes for  
CC measuring human gene expression in a sample derived from human heart (see  
CC ABA21535 ABA1305). The present sequence is a protein encoded by one such  
CC probe. The probes may be used for predicting, measuring and displaying  
CC gene expression in samples derived from the human heart via microarrays.  
CC By measuring gene expression, the probes are useful for predicting,  
CC diagnosing, grading, staging, monitoring and prognosing diseases of the  
CC human heart and vascular system e.g. cardiovascular disease,  
CC hypertension, cardiac arrhythmias and congenital heart disease.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 86 AA;

Query Match 9.7%; Score 437; DB 22; Length 86;  
Best Local Similarity 100.0%; Pred. No. 1.5e-24;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 374 GTSQSDISTLPVPSPPDLEVESETAMEVDTPAEQFLQPSSTMSQAQAHSTSSPTESPHS 433  
DB 1 GTSQSDISTLPVPSPPDLEVESETAMEVDTPAEQFLQPSSTMSQAQAHSTSSPTESPHS 60

OY 434 TPLISSPDSERQSVASGHHTHQS 459  
DB 61 TPLISSPDSERQSVASGHHTHQS 86

RESULT 27  
 AAM53809  
 ID AAM53809 standard; Protein; 86 AA.  
 XX  
 AC AAM53809;  
 XX  
 DT 05-NOV-2001 (first entry)  
 XX  
 DE Human brain expressed single exon probe encoded protein SEQ ID NO: 25914.  
 XX  
 KW Human: brain expressed exon; gene expression analysis; probe;  
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
 KW epilepsy; cancer.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157275-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US00667.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-483446/52.  
 XX  
 PT Single exon nucleic acid probes for analyzing gene expression in human  
 PT brains -  
 XX  
 PS Example 4; SEQ ID NO: 25914; 650bp + Sequence Listing; English.  
 XX  
 CC The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC brain. They can be used to measure gene expression in brain cell samples,  
 CC which may enable the diagnosis and improved treatment of nervous system  
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
 CC epilepsy and cancers. The present sequence is a protein encoded by one of  
 CC the probes of the invention.  
 XX  
 SQ Sequence 86 AA:  
 XX  
 Query Match 9.7%; Score 437; DB 22; Length 86;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-24;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 374 GTSQSDISTLPVPSSPDLEVSFTAMEVDTPAQFQIPSTSTMSQAASHSTSPSPHS 433  
 DB 1 GTSQSDISTLPVPSSPDLEVSFTAMEVDTPAQFQIPSTSTMSQAASHSTSPSPHS 60  
 OY 434 TPLSSPDSEQRQSVASGHTTHQS 459  
 DB 61 TPLSSPDSEQRQSVASGHTTHQS 86  
 RESULT 28  
 AAM6195  
 ID AAM6195 standard; Protein; 86 AA.  
 XX  
 AC AAM6195;  
 XX  
 DT 06-NOV-2001 (first entry)

XX  
 DE Human bone marrow expressed probe encoded protein SEQ ID NO: 26501.  
 XX  
 KW Human: bone marrow expressed exon; gene expression analysis; probe;  
 KW microarray; cancer; leukaemia; lymphoma; myeloma.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157276-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US00668.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-488900/53.  
 XX  
 PT Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human bone marrow -  
 XX  
 PS Example 4; SEQ ID NO: 26501; 658bp + Sequence Listing; English.  
 XX  
 CC The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC bone marrow. They can be used to measure gene expression in bone marrow  
 CC samples, which may enable the improved diagnosis and treatment of cancers  
 CC such as lymphoma, leukaemia and myeloma. The present sequence is a  
 CC protein encoded by one of the probes of the invention.  
 XX  
 SQ Sequence 86 AA:  
 XX  
 Query Match 9.7%; Score 437; DB 22; Length 86;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-24;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 374 GTSQSDISTLPVPSSPDLEVSFTAMEVDTPAQFQIPSTSTMSQAASHSTSPSPHS 433  
 DB 1 GTSQSDISTLPVPSSPDLEVSFTAMEVDTPAQFQIPSTSTMSQAASHSTSPSPHS 60  
 OY 434 TPLSSPDSEQRQSVASGHTTHQS 459  
 DB 61 TPLSSPDSEQRQSVASGHTTHQS 86  
 RESULT 29  
 AAM14066  
 ID AAM14066 standard; Protein; 86 AA.  
 XX  
 AC AAM14066;  
 XX  
 DT 12-OCT-2001 (first entry)  
 XX  
 DE Peptide #500 encoded by probe for measuring cervical gene expression.  
 DE Probe: human; microarray; gene expression; cervical epithelial cell;  
 KW cervical cancer.  
 KW  
 OS Homo sapiens.  
 XX  
 PN WO200157278-A2.  
 XX  
 PD 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00670.  
PE  
XX 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488901/53.  
XX  
XX  
XX Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human cervical epithelial cells -  
XX  
PS Claim 27; SEQ ID NO 18892; 487bp; English.  
XX  
XX The present invention relates to human single exon nucleic acid probes  
CC (SENPs; see A110068-A1128459). The present sequence is a peptide encoded  
CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs  
CC can be used to produce a single exon microarray, which can be used for  
CC measuring human gene expression in a sample derived from human cervical  
CC epithelial cells. By measuring gene expression, the probes are therefore  
CC useful in grading and/or staging of diseases of the cervix, notably  
CC cervical cancer.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [http://wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).  
XX  
SQ Sequence 86 AA;  
XX  
Query Match 9.7%; Score 437; DB 22; Length 86;  
Best Local Similarity 100.0%; Pred. NO. 1.5e-24;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 374 GTSQSDISTLPVSPSPDLEVSRTAMEVDYPAEOPLOPSTSTMSQAQHSSTSPRESPS 433  
DB 1 GTSQSDISTLPVSPSPDLEVSRTAMEVDYPAEOPLOPSTSTMSQAQHSSTSPRESPS 60  
XX  
QY 434 TPLISSPDSEQRQSVASGHHTHQS 459  
DB 61 TPLISSPDSEQRQSVASGHHTHQS 86  
XX  
RESULT 30  
AAM26472  
ID AAM26472 standard; Protein; 86 AA.  
XX  
AC AAM26472;  
XX  
DT 17-OCT-2001 (first entry)  
XX  
DE Peptide #509 encoded by probe for measuring placental gene expression.  
XX  
KW Probe: microarray; human; placenta; antenatal diagnosis;  
KM genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200157272-A2.  
XX  
PD 09-AUG-2001.  
XX  
PE 30-JAN-2001; 2001WO-US00663.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-48897/53.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human placenta -  
XX  
PS Claim 27; SEQ ID NO 26741; 654bp; English.  
XX  
XX The present invention relates to single exon nucleic acid probes (SENPs;  
CC see A113135-A1157546). The present sequence is a peptide encoded by one  
CC such probe. The probes are useful for producing a microarray for  
CC predicting, measuring and displaying gene expression in samples derived  
CC from human placenta. The probes are useful for antenatal diagnosis of  
CC human genetic disorders.  
XX  
SQ Sequence 86 AA;  
XX  
Query Match 9.7%; Score 437; DB 22; Length 86;  
Best Local Similarity 100.0%; Pred. NO. 1.5e-24;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 374 GTSQSDISTLPVSPSPDLEVSRTAMEVDYPAEOPLOPSTSTMSQAQHSSTSPRESPS 433  
DB 1 GTSQSDISTLPVSPSPDLEVSRTAMEVDYPAEOPLOPSTSTMSQAQHSSTSPRESPS 60  
XX  
QY 434 TPLISSPDSEQRQSVASGHHTHQS 459  
DB 61 TPLISSPDSEQRQSVASGHHTHQS 86  
XX  
RESULT 31  
AAM01806  
ID AAM01806 standard; Protein; 86 AA.  
XX  
AC AAM01806;  
XX  
DT 09-OCT-2001 (first entry)  
XX  
DE Peptide #488 encoded by probe for measuring human breast gene expression.  
XX  
KW Probe: human; breast disease; breast cancer; development disorder;  
KM inflammatory disease; proliferative breast disease; non-carcinoma tumour.  
XX  
OS Homo sapiens.  
XX  
PN WO200157270-A2.  
XX  
PD 09-AUG-2001.  
XX  
PE 29-JAN-2001; 2001WO-US00661.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-476286/51.  
XX

PT Novel single exon nucleic acid probe used to measuring gene expression  
 PT in a human breast -  
 XX  
 XX  
 PS Claim 27; SEQ ID No 10546; 322pp; English.  
 XX  
 CC The present invention relates to novel single exon nucleic acid probes  
 CC (see A1100010-A1110067). The present sequence is a peptide encoded by one  
 CC such probe. The probes are useful for measuring human gene expression in  
 CC a human breast sample, where the probe hybridizes at high stringency to a  
 CC nucleic acid expressed in the human breast. The probes are useful for  
 CC predicting, diagnosing, grading, staging, monitoring and prognosis  
 CC diseases of the human breast, particularly those diseases with polygenic  
 CC aetiology. The diseases include: breast cancer, disorders of development,  
 CC inflammatory diseases and non-carcinoma tumours.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX  
 SQ Sequence 86 AA:  
 9.7%; Score 437; DB 22; Length 86;  
 Query Match Best Local Similarity 100.0%; Pred. No. 1.5e-24;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 374 GTSQSDISTLPTVPSSPDLEVSSETAMEVDTPAEQFLQPSSTSMQAHSSTSPSPHS 433  
 DB 1 GTSQSDISTLPTVPSSPDLEVSSETAMEVDTPAEQFLQPSSTSMQAHSSTSPSPHS 60  
 OY 434 TPLSSPSEQRQSVESAGHHTHHOS 459  
 DB 61 TPLSSPSEQRQSVESAGHHTHHOS 86  
 RESULT 32  
 ABG35845  
 ID ABG35845 standard; Peptide; 86 AA.  
 XX  
 AC ABG35845;  
 XX  
 DT 19-AUG-2002 (first entry)  
 XX  
 DE Human peptide encoded by genome-derived single exon probe SEQ ID 25510.  
 XX  
 KW Human; single exon probe; asthma; lung cancer; COPD; ILD;  
 KW chronic obstructive pulmonary disease; interstitial lung disease;  
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
 KW tuberculous sclerosis; Gaucher's disease; Niemann-Pick disease;  
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagenen syndrome;  
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
 KW primary ciliary dyskinesia; pulmonary hypertension;  
 KW hyaline membrane disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200186003-A2.  
 PD 15-NOV-2001.  
 PD  
 PD 30-JAN-2001; 2001WO-US00665.  
 PF  
 XX 04-FEB-2000; 2000US-180312P.  
 PR 26-MAY-2000; 2000US-207456P.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-234687P.  
 PR 27-SEP-2000; 2000US-236359P.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 PA Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 PI

XX  
 DR WPI; 2002-114183/15.  
 XX  
 PT Spatially-addressable set of single exon nucleic acid probes, used to  
 PT measure gene expression in human lung samples -  
 XX  
 XX  
 PS Claim 27; SEQ ID No 25510; 634pp; English.  
 XX  
 CC The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human lung comprising single exon nucleic acid probes having one of  
 CC 12614 nucleic acid sequences mentioned in the specification, or their  
 CC complements or the 12387 open reading frames derived from the 12614  
 CC probes. Also included are a microarray comprising the novel set of  
 CC probes; the novel set of probes which hybridise at high stringency to a  
 CC nucleic acid expressed in the human lung; measuring gene expression in a  
 CC sample derived from human lung, comprising (a) contacting the array with  
 CC a collection of detectably labeled nucleic acids derived from human lung  
 CC mRNA, and (b) measuring the label detectably bound to each probe of  
 CC the array; identifying exons in a eukaryotic genome, comprising  
 CC (a) algorithmically predicting at least one exon from genomic sequences  
 CC of the eukaryote; and (b) detecting specific hybridisation of detectably  
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
 CC having a fragment identical to the predicted exon, the probe is included  
 CC in the above mentioned microarray; assigning exons to a single gene,  
 CC comprising (a) identifying exons from genomic sequence by the method  
 CC above and (b) measuring the expression of each of the exons in several  
 CC tissues and/or cell types using hybridisation to a single exon  
 CC microarrays having a probe with the exon, where a common pattern of  
 CC expression of the exons in the tissues and/or cell types indicates that  
 CC the exons should be assigned to a single gene; a peptide comprising one  
 CC of 12011 sequences, mentioned in the specification, or encoded by the  
 CC probes/open reading frames (ORF). The probes are used for gene  
 CC expression analysis, and for identifying exons in a gene, particularly  
 CC using human lung derived mRNA and for the study of lung diseases  
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease  
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary  
 CC fibrosis, neurofibromatosis, tuberculous sclerosis, Gaucher's disease,  
 CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary  
 CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,  
 CC pulmonary alveolar proteinosis, Karagenen syndrome, fibrocystic  
 CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension  
 CC and hyaline membrane disease. The present sequence is a peptide/protein  
 CC encoded by a single exon probe of the invention.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX  
 SQ Sequence 86 AA:  
 9.7%; Score 437; DB 23; Length 86;  
 Query Match Best Local Similarity 100.0%; Pred. No. 1.5e-24;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 374 GTSQSDISTLPTVPSSPDLEVSSETAMEVDTPAEQFLQPSSTSMQAHSSTSPSPHS 433  
 DB 1 GTSQSDISTLPTVPSSPDLEVSSETAMEVDTPAEQFLQPSSTSMQAHSSTSPSPHS 60  
 OY 434 TPLSSPSEQRQSVESAGHHTHHOS 459  
 DB 61 TPLSSPSEQRQSVESAGHHTHHOS 86  
 RESULT 33  
 AAG18387  
 ID AAG18387 standard; Protein; 471 AA.  
 XX  
 AC AAG18387;  
 XX  
 DT 17-OCT-2000 (first entry)  
 XX  
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 19771.

XX Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridization assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
OS Arabidopsis thaliana.  
PN EP1033405-A2.  
XX  
XX 06-SEP-2000.  
PD  
XX 25-FEB-2000; 2000EP-0301439.  
PF  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145224.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145216.  
PR 27-JUL-1999; 99US-0145591.  
PR 27-JUL-1999; 99US-0145591.  
PR 27-JUL-1999; 99US-0145591.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147182.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.

```

PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152263.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159337.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 9.6%; Score 436.5; DB 21; Length 471;

Best Local Similarity 18.7%; Pred. No. 2,2e-23;

Matches 147; Conservative 66; Mismatches 153; Indels 419; Gaps 12;

```

QY 9 HLLMDVRRKSLGLEDPSRLSRYLGRREFIQRKLEATLVNVDGCVNTICMNDGGEYILS 68
DB 13 HSYVWVWERELGLNRSFNRSFASSEDLLRLGLDKLDRKKGCVNYVSPFADGDILIS 72
QY 69 GSDPKIVISNDYSRKRYLTTRSGHRANIPSAKFLPTCNDKOIVSCSGDGVITYNVED 128
DB 73 GSDRQVILMWMQATSVLSPDSGFHNNIFQAKFMPFSDRITVNSADKQVRSKILS 132
QY 129 AETNQCQFOTCHYGTTEIMVYPNDPYTELSCGEGDGYRWPTRIKTSCTKDKDDILL 188
DB 133 GOVEVSL-LGRHQGVYHLAVPEPSFSPYTCGEDGAVKHFELRTVATNLETC-EAKF 190
QY 189 NCRRAATVATCPPIPYLLAVGCSDSVRIYDRMLGTATGNVYGRGTGVARFIPSH 248
DB 191 NLVYVLIHALADPNRPGLLAVAGNDEYARVDIRYRESEGNWF-----TOPIDHFCGH 245
QY 249 L-NKSCSVTSLCYSEDQELIVSYSDYIYLPDKDDTARLKPASAEERREELROPV 307
DB 246 LIGDHWGITGLAFS-DOSELLASYSDEFIYTFP----- 279
QY 308 KRLRLRGWMTGPRARERERENDGEGSPNVSLMQRMSSDMLSRWFEEASEVAQSNRG 367

```

```

DB 280 -----DM-----GLG 284
QY 368 KSRPRGRTSOSDISTLPTVPSSPDLEVSETAMEVDTPAEQFLQPSSTSTMSQAQAHSTSSP 427
DB 285 -----P 285
QY 428 TESPHSTPLSPDSRQSVASGHHTHQSDNNNEKLSPKGTGPVLSLHSTGTT 487
DB 286 TPYPSSTK----- 293
QY 488 TSTIKLFTDEWSSIASSSNGISGHCKSEGOEESFVPOSSVQPPGDSFKAPEESSEDV 547
DB 294 -----TERMPPO----- 301
QY 548 TKYQGVSAENPVENHINITQSDKFTAKPLDINSGERNDLINDRSCGVPEASSEKAKE 607
DB 302 ----- 301
QY 608 PETSDQSTESATENNNTNPEQFQTEATGPAHEETSTRDSALQDTDDSDDPVILPGA 667
DB 302 ----- 301
QY 668 RYRAGDRRSAVARIQEFFFFRKRKEMEBELDTNIRPLVMMYKGRHSRTMIKEAN 727
DB 302 -----YKEHTNRET-YKGVN 316
QY 728 FNG--ANFYMSGSDGHIITMDRHTAHLMLADNHYVNCLOPHFPDPLIASSGIDYDI 785
DB 317 FPGKCEYVAGSDGCRIFIRWKRKDELRLAMEADRHVNCIESHBMPLMCSGIDTDI 376
QY 786 KIWSF 790
DB 377 KIWTF 381

```

RESULT 34

AAG50441

ID AAG50441 standard; Protein; 471 AA.

XX AAG50441;

XX 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 63925.

KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130049.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132488.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140693.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.

PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148566.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 23-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151080.  
PR 27-AUG-1999; 99US-0151303.  
PR 30-AUG-1999; 99US-0151348.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0152363.  
PR 07-SEP-1999; 99US-0153070.  
PR 10-SEP-1999; 99US-0153758.  
PR 13-SEP-1999; 99US-0154018.  
PR 15-SEP-1999; 99US-0154039.  
PR 16-SEP-1999; 99US-0154779.  
PR 20-SEP-1999; 99US-0155139.  
PR 22-SEP-1999; 99US-0155486.  
PR 23-SEP-1999; 99US-0155659.  
PR 24-SEP-1999; 99US-0156458.  
PR 28-SEP-1999; 99US-0156596.  
PR 29-SEP-1999; 99US-0157117.  
PR 04-OCT-1999; 99US-0157753.  
PR 05-OCT-1999; 99US-0157865.  
PR 06-OCT-1999; 99US-0158029.  
PR 07-OCT-1999; 99US-0158232.  
PR 08-OCT-1999; 99US-0158368.  
PR 12-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.



```

PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160881.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 9.6%; Score 436.5; DB 21; Length 471;  
 Best Local Similarity 18.7%; Pred. No. 2.2e-23;  
 Matches 147; Conservative 66; Mismatches 153; Indels 419; Gaps 12;

```

QY 9 HILMDVYKRSGLGDEPSRLSRVIGRREFLORLKLATLVNHDGCYNTICMNDTGEYILS 68
DB 13 HAVVWVWERELGLPRRSNSRPSASEDLRLGLDKLDRKGCYNTVSPFNADGILLS 72
QY 69 GSDDTKLIVSNPYSRKVLTTINSGRANIFSAKFLPCTNDKOIVSCSGDGVIFYTVEOD 128
DB 73 GSDDBQVILMDQVIAVSKLISFDSGHNIFQAKFMFSDRITVSAADKQVRSKILMS 132
QY 129 AETNRQCOFTCHYGTTEIMTVPNDDPYTFLSCGEGTVMFPRITKSTCKEDCKDII 188
DB 133 GOVERSL-LGRKQGVHKLAVEPGSPFSTYTGEDGAVAHFDRKVAATNLFCK-EAKF 190
QY 189 NCRRAATVAICPIPIYVLAVGSCDSVRIYDRMLGTRATNGVAGRTGAVARFIPSH 248
DB 191 NLVYVLIHAIADPDRMGLLAAGMDEVAVYDIRSRSQWNF-----TQIDHRCPEH 245
QY 249 L-NNNSCAVTLSCYSEDQDEILVSYSSDIYIFLPDKDPTARLKTSPASERREELQPPV 307
DB 246 LIGDHVIGITGLAFS-DSSELLASYDEIFYLFTP----- 279
QY 308 KRLRLRGWSDYTPRPARPESERERDGEOSPVSILQMSDMLSRWFEASEVAQSNRG 367
DB 280 -----DM-----GLG 284
QY 368 RSRPRGTSOSDITLPTVPSSPDLEVSETAMEVDTPAEQFLQPSSTMSQAHSSTSP 427
DB 285 -----P 285
QY 428 TESPSPRLSSPDSEQSQVEASGHTHHQSDNNNEKLSKPGICEVYLISHSTEGTT 487
DB 286 TPYSST----- 293
QY 488 TSTIKLFTDEWSSIASSSRGISGCKSEGOEESFVPOSSVOPPEGDEFTKAPERSSEV 547
DB 294 -----TEERTPQ----- 301
QY 548 TKYQEGVSAENPVENHNIITQSDKFTAKPLDSSNGERNLMDNSCGVPEESASSEAKE 607
DB 302 ----- 301
QY 608 PETSDQSTESATNENNTNPEPOTATGPAHETSTRSALQDDDDDDVILPGA 667
DB 302 ----- 301
QY 668 RYRAGPDRSAAVARIQEFFRRKREKEMELDTLNRPLVYKMYGKGRNSRTMIKEAN 727
DB 302 -----VYKEHNREI-VAGVN 316
QY 728 FWG--ANFVWSGSDCGHIFIMDRHTAEHMLLEADNHVVCLOPHPPDILASSGDIYDI 785

```

```

DB 317 FFGPKCEVWVSGSDCGRIFTRKKGDELRAHEADRHVNCIESHPHPLMCSGIDTID 376
QY 786 KIMSP 790
DB 377 KIMTP 381

```

## RESULT 35

AA85870 standard; Protein; 816 AA.

AA85870;

13-SEP-1996 (first entry)

WD-40 domain-contg. Mus musculus protein.

WD40 repeat region; beta-transducin; protein-protein interaction; drug; intracellular signalling; protein kinase C; homology; motif; modulator; receptors of activated protein kinase; enzyme activity; isozyme; human.

Mus musculus.

WO9521252-A2.

10-AUG-1995.

31-JAN-1995; 95WO-US01210.

01-FEB-1994; 94US-0190802.

(STRD ) UNIV IELAND STANFORD JUNIOR.

Mochly-Rosen D, Ron D;

WPI; 1995-283772/37.

New WD-40 (beta-transducin)-derived polypeptide(s) - which alter the activity of a protein, eg. protein kinase C, which interacts with a protein contg. a WD-40 region.

Example 5; Page 135-138; 351pp; English.

proteins AAR8581-92 are protein which contain at least one WD-40 (also called beta-transducin homologous) amino acid repeat motifs. The WD-40 regions are involved in protein-protein interactions between proteins involved in intracellular signalling. An example of such an interaction is between protein kinase C and receptors of activated protein kinase (RACK), esp. RACK-1 (AAR85850). Proteins AAR8581-82 were isolated based on homology with beta-transducin, whereas proteins AAR8582-92 were isolated based on homology with the WD-40 consensus sequence (AAR85893). The proteins were used to construct the peptides AAR84928-85063 and AAR85786-85842. The peptides can be used to identify target proteins contg. WD-40 motifs, as modulators of enzyme esp. isozyme, activity of proteins involved in protein-protein interaction and to screen for drugs that will affect protein-protein interaction involving WD-40 domains.

Sequence 816 AA;

Query Match 9.5%; Score 429.5; DB 16; Length 816;  
 Best Local Similarity 18.7%; Pred. No. 1.7e-22;  
 Matches 164; Conservative 76; Mismatches 173; Indels 465; Gaps 18;

```

QY 8 PHILMDV-----RKRSLGDEPSRLSRVIGRREFLORLKLATLVNHDGCYNTICMNDTG 63
DB 377 PRPRMNVLSALRDQGLG--SSGRFYVEACGARKLVQRFSLHVEGSGCVNTHFMDHG 434
QY 64 EYILSGSDTKLIVSNPYSRKVLTTINSGRANIFSAKFLPCTNDKOIVSCSGDGVIFYT 123
DB 435 TLIASGSDDLKIVIMDWLKRSLVNFDSGHNILQAKFLPNCNDAILAMGRCQGVARA 494
QY 124 NVEQDAETNQCQFTCHYGTTEIMTVPNDDPYTFLSCGEGTVMFPRITKSTCKEDCK 183

```

```

DB 495 QLSAVACTHTKRLVKGASHRLGLEPDSFRRLTSGEDAVVEFIDLRQAHPAK----- 550
OY 184 DDILI-----NCRRAATSAICPPIPYLLAVGSSSVRIYDRMLGTRATGVAAGRTTG 239
DB 551 --LLVINDGSKKVLTYFVFPANVYGFVAGGDDFMRIYDQRKIDENV-----NNG 600
OY 240 MVAEFTSHL--NNSCRVTSLSYSEGOELIVYSDDYITLFDPKDPAELKTPSAE 297
DB 601 VLKFCFPHLLSDYPAPAHITSLMYSDDELLASTYNDDEITYE----- 643
OY 298 RRELQDPVYKRLRLRGDMDTPRARPERERDEQSPNVSLQMSDMLRWEAS 357
DB 644 ----- 643
OY 358 EVAQSNRGRSRPRGTSQSDISTLPTVPSSDLEVESETAMEVDTPAQFLQPTSSIM 417
DB 644 ----- 643
OY 418 SQAQHTSPTSPHSTLLSSPDSEQRQSVASGHHTHQSDNNNEKLSPKGTGPVL 477
DB 644 ----- 643
OY 478 SLHSTEGTTSTIKLFTDEWSIASSSRGIGSHCKSEGOESFVQSSVOPPEGDSST 537
DB 644 ----- 643
OY 538 KAPESSBDVTKYQBGVSAENPVENHINITQSDKFLAKPLDSSGGRNDLNDRSGVPE 597
DB 644 ----- 643
OY 598 ESASSEKAKEPETSQSTESATNENNTNPEQFQTEATGSAHESTSPDSALQDTDS 657
DB 644 -----NSDS 648
OY 658 DDDPYLIPGARYRAGGDRRSAAVARIQEFFRRRKEKMEELDTLNRPLVAKMYGHR 717
DB 649 D-----GAQY-----AKR-----YKGR 661
OY 718 NSRTMIKEANFWG--ANFVMSGSDGHIPTDRHTAHLMLLEAD-NHVVNCLQPHFDP 774
DB 662 NNST-VKGVYFYGPNSSEFVMSGSDCGHIFTWKSQIOYQFLDEBEGTINCIDSHYLP 720
OY 775 ILASSGIDYDIKISPLESRIFNRKLA--DEVITRNEMLER--TNTITVPAS---FM 827
DB 721 VLASSGLDHEVKIWSPIAE--SKLAGLKNYIKIKLRDNFTLRTSTLFSNMLCFL 777
OY 828 LRMASLNLH-----IRADRLBGDRSGSGQENENED 859
DB 778 MSHVTSQSNYGRSWRGIRINAGGDFSDSSSSSEETNOE 815

RESULT 36
ABG27224
ID ABG27224 standard; protein; 633 AA.
AC ABG27224;
XX
XX
XX 18-FEB-2002 (first entry)
DE Novel human diagnostic protein #27215.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX

```

```

PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA Drmanac RT, Liu C, Tang YT.
PI
XX
XX MPI; 2001-639362/3.
DR N-PSDB; AAS91411.
XX
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensic, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 20; SEQ ID NO 57583; 103pp; English.
XX

```

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome CC and gene mapping, and in recombinant production of (II). The CC polynucleotides are also used in diagnostics as expressed sequence tags CC for identifying expressed genes. (I) is useful in gene therapy techniques CC to restore normal activity of (II) or to treat disease states involving CC (II). (II) is useful for generating antibodies against it, detecting or CC quantitating a polypeptide in tissue, as molecular weight markers and as CC a food supplement. (II) and its binding partners are useful in medical CC imaging of sites expressing (II). (I) and (II) are useful for treating CC disorders involving aberrant protein expression or biological activity. CC The polypeptide and polynucleotide sequences have applications in CC diagnostics, forensic, gene mapping, identification of mutations CC responsible for genetic disorders or other traits to assess biodiversity CC and to produce other types of data and products dependent on DNA and CC amino acid sequences. ABG00010-ABG30377 represent novel human CC diagnostic amino acid sequences of the invention. CC Note: The sequence data for this patent did not appear in the printed CC specification, but was obtained in electronic format directly from WIPO CC at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 633 AA;

Query Match 8.9%; Score 404; DB 22; Length 633;

Best Local Similarity 17.9%; Pred. No. 8.5e-21;

Matches 160; Conservative 81; Mismatches 168; Indels 486; Gaps 21;

```

OY 2 SRGGSTPHLLMDV-----RRSLGLEPDSRLRSRY--LGRREFLQRLKLEATLWHDGCYN 55
DB 179 SETSALPRPMOVYVTLHORQIG---SRREYVACGARAFOKRFLOLRADHVCYN 234
OY 56 TICMNDTGXYILSGSDTKLIVSNPYSRKVLTTIRSGHRANISAKFLPCTNDKQIVSCS 115
DB 235 TVHFNQGRTRILASSGDDKIVYVMDWRQRYLNFESGHTNNVQAKFLPMCGDSTLAMCA 294
OY 116 GDGVIFYTNEODAEETNRQOCFTCHYGTTEIMTVNDPTFTLSCGEDGYRWMDTR--- 172
DB 295 RDGOVVAELINASYFNNTKCVAQHNRGPAKRLLEPSPKFLTSGDAVFTIDLRQDR 354
OY 173 --IKTSCTEKEDCKDDILINORRAATSAICPPIPYLLAVGSSSVRIYDRMLGTRATG 230
DB 355 PASKVVVATRENDK-----KVLGYTTVPANTYQAVAGGDDFVRIYDQRKIDKE-- 405
OY 231 NYAGRTGKVAEFTSHLNNKSC---RYSLSYSEGOELIVYSDDYITLFDPKDPT 286
DB 406 -----NNGVLKFFFTPHLVN--CDFPTNITCVYSHDGETLLASTYNDDEITYE----- 452
OY 287 ARELKTPSAEERBELQDPVYKRLRLRGDMDTPRARPERERDEQSPNVSLQMS 346
DB 453 ----- 452
OY 347 DMLSRWFEASVAQSNRGRSRPRGTSQSDISTLPTVPSSDLEVESETAMEVDTPAE 406
DB 453 ----- 452

```

```

QY 407 QELPSTSTMSQAHSSTSPSPSTPLSSPSEORQSVASGHTHQSDDNNKTL 466
Db 453 -----
QY 467 SPKPGTGPVLIAHSTEGTSTIKLNPTEWSSIASSSRGISGHSCKSEGESEFVQS 452
Db 453 -----
QY 527 SVQPPGDSSEKAPESSEEDVTKOEGVSAENPVENHINTOSDKFTAKPLDSNGERN 452
Db 453 -----
QY 587 LINDRCGVPEASASEKAPETSDQSTESATNENTNPEPOQTATGSAHEETST 452
Db 453 -----
QY 647 RDSALQDTDDSDPVLIGARYRAGPGRRSAAVARIOEFFRRRERKEMELDTLNR 457
Db 458 -----
QY 707 PLVKNVKGHRNRTMKRANFNG--ANFVMSGSDCGHFTWDRHTAEHLMLLEADNH-V 464
Db 465 -----
QY 764 VNCIQHPPEPIIASGIDYDIKIMSPLESRIFFRKLADN-----VITREML--EET 517
Db 518 INCLEPHPLPYLACSGLDHDKIWP-----TAKAFELTGLKVKYIKKNMERDEDS 570
QY 817 RNTTVASFMLRLASLNH-----RADRLP--GDRSGSGGENENEDDE 860
Db 571 LHHGSLFDQYMLWFL--LRHYTORGRHQDWRSGEAFPPDESDSSSTSESEE 623

RESULT 37
AAU80461
ID AAU80461 standard; Protein; 385 AA.
AC AAU80461;
DE 12-MAR-2002 (first entry)
DE Fruit fly adipose protein, adp, #1.
KW Adipose protein; adp; obesity; transgenic animal; obesity;
KW adipositas; bulimia; wasting; cachexia; eating disorder;
KW body weight disorder; weight loss; cancer; infectious disease;
KW hypogonadism; Prader-Willi syndrome; Laurence-Moon-Biedl syndrome;
KW hypothyroidism; diabetes; Cushing's syndrome; endocrine disorder;
KW gastrointestinal diseases; inflammatory bowel disease;
KW ulcerative colitis; anorexia nervosa; glycogen storage disease;
KW lipid storage disease; lipoma; liposarcoma; heart disease; hypertension;
KW infertility; acquired immunodeficiency syndrome; AIDS.
OS Drosophila melanogaster.
XX
XX WO200196371-A2.
XX
XX 20-DEC-2001.
XX
XX 13-JUN-2001; 2001WO-EP06713.
XX
XX 16-JUN-2000; 2000US-211914P.
XX 23-JUN-2000; 2000EP-0113049.
XX 28-JUN-2000; 2000US-214518P.
XX 17-APR-2001; 2001EP-0109537.
XX
XX (DEVE-) DEVELOGEN AG.
XX
XX Brocner G, Ciossek T, Dohrmann C, Haeder T, Rothe M;
XX WPI; 2002-106464/14.
XX N-PSDB; ABK16344.
XX

```

PT Novel nucleic acid encoding adipose polypeptide which regulates, causes  
PT or contributes to obesity, useful for treating obesity, heart disease,  
PT hypertension, infertility, and controlling weight loss in cancer  
PS patients -  
XX  
XX Claim 1: Page 138-139; 188pp; English.

The invention relates to a nucleic acid encoding a adipose (ADP)  
CC polypeptide which regulates, causes or contributes to obesity in an  
CC animal or a human. The polynucleotides, proteins, anti-adp antibodies,  
CC vectors, adp activity, adp antisense nucleic acids, expression  
CC treatment of obesity, transgenic animals are useful in the diagnosis and  
CC disorders and/or disorders of body weight/body mass, weight loss due to  
CC cancer or infectious diseases, genetic disorders associated with  
CC hypogonadism e.g. Prader-Willi syndrome, Laurence-Moon-Biedl syndrome,  
CC hypothyroidism, diabetes, Cushing's syndrome, endocrine disorders,  
CC colitis, and anorexia nervosa. They are also useful for treating  
CC disorders of body weight/mass e.g. glycogen storage diseases, and lipid  
CC storage diseases and for treating lipomas, and/or liposarcomas. The  
CC compositions are also useful for treating heart disease, hypertension,  
CC e.g. enhancing or controlling fertility, controlling weight loss in  
CC acquired immunodeficiency syndrome (AIDS) or cancer patients. The  
CC present sequence is an adp protein.

Sequence 385 AA;

Query Match 8.88; Score 399; DB 23; Length 385;  
Best Local Similarity 32.6%; Pred. No. 9.3e-21;  
Matches 93; Conservative 50; Mismatches 120; Indels 22; Gaps 5;

```

QY 10 LMDVRRKSLGLDPSRLSRRLGRREFIQRLKLEALNVHDCVNTICWMDGEYILSG 69
Db 28 LHM--QROQYGHLEDEMLRLRLASPAYVDRLDEALVHGEGVNCLETTDGMASG 85
QY 70 SDDTKLVISNPSRKVLITIRSGHRANIFSAKFLPCYNDKQIVSCSGDGVIEFTNVQDA 129
Db 86 SDDRYVMWMDPFRKKLVHVRITKHLGVFVKLPKNNISIVATCADKTIYYVDNDPN 145
QY 130 ETRNCOFTCHYTYEIMKVPNDPTFLSCGEDGVYRWMDTRIKTSCTEDCKDILLIN 189
Db 146 ETLFSC--ICHFSRANRLAQAQSPHYFWSAGDGLQIDIREPHRCRREGIVRLIN 203
QY 190 C-----RRATSVAILCPPIPYLAVGCSQSVRYVDRMIGSTRATGNYAGRTTGMVYAR 244
Db 204 LHDQLENTAKCLAINPRREYLAAGTNDPFARVYDRKKLPST-----NNGISACVAYY 258
QY 245 IRSHLNKNSCR-----VTSICYSDEQGLIVSYSSDYLYFD 281
Db 259 APQIVKNISRNIVHEPRGITVLTENGCTELLVNICGEHYVAFD 303

```

```

RESULT 38
ABB57847
ID ABB57847 standard; Protein; 748 AA.
XX
XX ABB57847;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 333.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX

```

PF 23-MAR-2001: 2001MO-US09231.  
 XX 23-MAR-2000: 2000US-191637P.  
 PR 11-JUL-2000: 2000US-0614150.  
 XX  
 PA (PEKE) PE CORP NY.  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX WPI; 2001-656860/75.  
 DR N-PSDB; ABL01950.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 PS  
 XX Disclosure; SEQ ID NO 333; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC capable in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
 CC sequences (AB161737-AB16175) and the encoded proteins  
 CC (AB161737-AB16175).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX Sequence 748 AA:

Query Match 8.2%; Score 373; DB 22; Length 748;  
 Best Local Similarity 17.3%; Pred. No. 2.1e-18;  
 Matches 147; Conservative 73; Mismatches 163; Indels 468; Gaps 14;

QY 31 YLGRREFIOLKLEATNVGDCVNTICWMDTGEYILSGSDTKLVISNPKRYLTTIR 90  
 DB 306 YYSROVVEOMTLISLNVHGCYNSLNFNAGDLICSGSDLTIVVDMWMEKOLHFR 365  
 QY 91 SGRHANIFSAKFLPCTNDKOTIVSCSGDGVIFVTVEODAEFNRCQCFCHTTEIMV 150  
 DB 366 SGHNMMIFQFKFIDSGCLDIVSSRDQVRSYVPPSGVYKPIRLTYHSESVKILIV 425  
 QY 151 PNDYTLSCGEDGTVMFPRITKSCYKEDOKDILINC-----RRATSYALCP 201  
 DB 426 PHSRHEILSAGEDAAYVHFELASNAAT-----TMRQVNDENRERVALFSTAHNP 478  
 QY 202 PIPYLLAVGSDSVIRIDRRMLGTRATGNVAGRGTTGNAVAFIPSHLNKSCRYSLCY 261  
 DB 479 YAFPCVSG-SDILIRVYDKRL-AKAIHQMAPRNL-----LEAOITQITCAVY 525  
 QY 262 SEDQELIVSYSSDIYTFDPKODTARELKTSAERREBELQPPYKRLRLRGDMSDTGP 321  
 DB 526 NHSGEILIASYSDAGITLFDSENYN-----RDY----- 554  
 QY 322 RARPERERBERGEGSPNVSLQMSDMLSRWESEVAQSNRGSRPRGGTSGSDIS 381  
 DB 555 ----- 554  
 QY 382 TLPVPSPDLEVSETAMEVDTPAEOFLQPTSTSTMSAQAHSTSPESHSTPLSSPD 441  
 DB 555 ----- 554  
 QY 442 SGRQROVEASGHTTHQSDNNNEKLSPPQGTGPVLISLHSTGTGTTSTIKLNFTEWSS 501  
 DB 555 ----- 554  
 QY 502 IASSRGISGSHCKSEGOESFVPOSSVQPEGSETPAKPESSEDYTKYQGVSAENPYE 561  
 DB 555 -----LHC----- 557  
 QY 562 NHINTQSDKFTAKPLDINSGERNDLDRSCGVPEESASSEKAKETSDOTSTESATN 621

DB 558 ----- 557  
 QY 622 ENNTNEPQOTATGSAHEETSTRDSALQDTDDSDDPVILPGARYRAGPGDRRSVA 681  
 DB 558 ----- 557  
 QY 682 RIQEFFRRKREKEMEELDTLIRRLPVKMYKGRNSRIMKEANFWG--ANTVMSSGD 739  
 DB 558 -----YKGHINSRT-IKGVNFPGRSEYIVSGSD 585  
 QY 740 CGHIFWDRHTAHEMLLEADN-HVNCLOPHPDPLASSGIDYIKIMSPLEESRIFN 798  
 DB 586 CGNIFWDMKTEALINMKGDHGVNCLPEHPMPVLAISLEDVAKITPQGERKLP 645  
 QY 799 RKLADETTRN-----ELMLEETNTI-----TVPASFMLMLASLNHRADRL 843  
 DB 646 EDLLKQTLQKRNFRNIVESIDLDINNFYIFRIGFQSGASHLRQHTSLSLGHQMDN-S 704

RESULT 39  
 ABB27841  
 ID ABB27841 standard; Peptide; 62 AA.

ABB27841;

01-FEB-2002 (first entry)

Human peptide #492 encoded by breast cell single exon nucleic acid probe.  
 Human; microarray; single exon probe; gene expression; breast;  
 disease; cancer.

Homo sapiens.

WO200157271-A2.

09-AUG-2001.

30-JAN-2001: 2001MO-US00662.

04-FEB-2000: 2000US-0180312.

26-MAY-2000: 2000US-0207456.

30-JUN-2000: 2000US-0608408.

03-AUG-2000: 2000US-0632366.

21-SEP-2000: 2000US-0234687.

27-SEP-2000: 2000US-0236359.

04-OCT-2000: 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Peen SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-496933/54.

New spatially-addressable set of single exon nucleic acid probes,  
 useful for measuring gene expression in sample derived from human  
 breast, comprises number of single exon nucleic acid probes -  
 Claim 27; SEQ ID NO 10809; 327pp + sequence listing; English.

The invention relates to a spatially-addressable set of single exon  
 nucleic acid probes for measuring gene expression in a sample derived  
 from human breast and for 474 cells. The method involves contacting  
 the probes with a collection of detectably labelled nucleic acids  
 derived from mRNA of human breast, and then measuring the label  
 bound to each probe of the microarray. The probes are useful for  
 verifying the expression of regions of genomic DNA predicted to  
 encode proteins. They are useful for gene discovery, and for

CC determining predisposition and/or prognosing breast disease. Gene  
 CC expression analysis is useful for assessing the toxicity of chemical  
 CC agents on cells. The microarray of this invention presents a far greater  
 CC diversity of probes for measuring gene expression, with far less bias  
 CC than expressed sequence tag microarrays. The method is suitable for  
 CC rapid production of functional information from genomic sequence. The  
 CC present sequence is a peptide encoded by a single exon nucleic acid  
 CC probe of the invention.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 SQ Sequence 62 AA;

Query Match  
 Best Local Similarity 100.0%; Score 338; DB 22; Length 62;  
 Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 85 VLTIRSGHRANIFSAKFLPCTNDKQIVSCSGDGVIFYTNVEDAETNRCCOFTCHYGT 144  
 DB 1 VLTIRSGHRANIFSAKFLPCTNDKQIVSCSGDGVIFYTNVEDAETNRCCOFTCHYGT 60  
 OY 145 YE 146  
 DB 61 YE 62

## RESULT 40

ABB33012  
 ID ABB33012 standard; Peptide: 62 AA.

AC ABB33012;

DT 04-FEB-2002 (first entry)

DE Peptide #518 encoded by human foetal liver single exon probe.

KM Human; foetal liver; gene expression; single exon nucleic acid probe.

OS Homo sapiens.

PN WO200157277-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00669.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-483447/52.

PT Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human fetal liver -

PS Claim 27; SEQ ID NO 25647; 639pp + sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for

CC measuring human gene expression in a sample derived from human foetal

CC liver. The single exon nucleic acid probes may be used for predicting,

CC measuring and displaying gene expression in samples derived from human

CC foetal liver. The present sequence is a peptide encoded by a single exon

CC nucleic acid probe of the invention.

CC Note: The sequence data for this patent did not form part of the

CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 SQ Sequence 62 AA;

Query Match  
 Best Local Similarity 100.0%; Score 338; DB 22; Length 62;  
 Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 85 VLTIRSGHRANIFSAKFLPCTNDKQIVSCSGDGVIFYTNVEDAETNRCCOFTCHYGT 144  
 DB 1 VLTIRSGHRANIFSAKFLPCTNDKQIVSCSGDGVIFYTNVEDAETNRCCOFTCHYGT 60  
 OY 145 YE 146  
 DB 61 YE 62

Search completed: March 18, 2003, 15:34:01  
 Job time : 54 secs

GenCore version 5.1.4-p5-4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 18, 2003, 15:30:15 ; Search time 37 Seconds

(without alignments)  
4789.203 Million cell updates/sec

Title: US-09-781-693a-2

Sequence: 1 MSRGSRPHLMVDKPKRLG.....RLEGRSRGSGQENEDDEE 860

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 segs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 120 summaries

Database :

SPTREMBL 21:  
1: sp\_archaea:  
2: sp\_bacteria:  
3: sp\_fungi:  
4: sp\_human:  
5: sp\_invertebrate:  
6: sp\_mammal:  
7: sp\_mhc:  
8: sp\_organelle:  
9: sp\_phage:  
10: sp\_plant:  
11: sp\_rodent:  
12: sp\_virus:  
13: sp\_vertebrate:  
14: sp\_unclassified:  
15: sp\_rviro:  
16: sp\_bacteriaphage:  
17: sp\_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3972	87.7	876	11	Q9DC22
2	2663	58.8	513	4	Q8TB19
3	2647	58.5	533	4	Q9P000
4	2258	49.9	493	10	Q9UD05
5	590.5	13.0	754	10	Q9LEU6
6	580	12.8	757	10	Q9AB03
7	547.5	12.1	809	3	Q94527
8	508.5	11.2	628	5	Q9V8C7
9	448.5	9.9	481	10	Q9M1E5
10	443.5	9.8	597	4	Q12839
11	429.5	9.5	747	11	Q01078
12	405.5	9.0	493	10	Q9U001
13	384.5	8.5	426	10	Q9S2M8
14	373	8.2	743	5	Q8SX92
15	373	8.2	748	5	Q9W091
16	330	7.3	615	4	Q9NV87

17	249.5	5.5	1514	5	Q8SY55
18	245.5	5.4	355	4	Q9HBE2
19	232.5	5.1	958	4	Q96JK2
20	213.5	4.7	747	5	Q9V517
21	212.5	4.7	298	5	Q95TD6
22	211.5	4.7	747	5	Q9B115
23	207	4.6	131	10	Q9FWE0
24	204.5	4.5	1217	4	Q60336
25	202	4.5	5170	5	Q17490
26	201	4.4	2586	5	Q9VTK8
27	200	4.4	1183	5	Q960U6
28	200	4.4	2351	5	Q9V054
29	200	4.4	6994	5	Q17343
30	199	4.4	493	5	Q23330
31	198	4.4	3507	5	Q23587
32	195.5	4.3	791	13	Q9PGL1
33	195.5	4.3	1162	5	Q9VB40
34	194.5	4.3	3060	5	Q9VAV4
35	194	4.3	1680	5	Q9UAO2
36	193	4.3	451	3	Q9P785
37	193	4.3	852	10	Q9S255
38	192.5	4.3	985	12	Q67643
39	191.5	4.2	2251	5	Q9W320
40	191	4.2	1508	11	Q9R010
41	190	4.2	788	4	Q96A69
42	189	4.2	819	4	Q96ST2
43	189	4.2	1142	4	Q8RCV4
44	189	4.2	2174	4	Q8TRT0
45	188.5	4.2	2174	5	Q9G0R0
46	187.5	4.1	2004	16	Q970P7
47	186.5	4.1	2283	2	Q8V099
48	186	4.1	1165	4	Q95819
49	185.5	4.1	1601	5	Q8WSM1
50	184	4.1	474	5	Q8R1B6
51	184	4.1	1255	2	Q86167
52	184	4.1	5722	5	Q9VPL9
53	184	4.1	5322	5	Q9N164
54	183	4.0	727	5	Q18918
55	183	4.0	2271	16	Q990Y4
56	183	4.0	2297	3	Q9HGK6
57	180.5	4.0	1175	4	Q75172
58	180.5	4.0	2910	10	Q9FNDS
59	178	3.9	1027	4	Q9NST7
60	178	3.9	1051	10	Q9LXN4
61	177.5	3.9	1063	16	Q9CH86
62	177.5	3.9	1100	13	Q57576
63	176.5	3.9	802	5	Q19764
64	176.5	3.9	5327	5	Q76891
65	176	3.9	703	13	Q90YE6
66	176	3.9	1459	5	Q17084
67	176	3.9	1838	13	Q9DC88
68	175.5	3.9	2768	5	Q9VC00
69	175.5	3.9	4900	5	Q9N541
70	175	3.9	534	3	Q94317
71	175	3.9	1604	3	Q21418
72	175	3.9	1829	5	Q22248
73	174.5	3.9	734	5	Q8T1A0
74	174.5	3.9	1587	10	Q9C8N3
75	174	3.8	1200	5	Q25388
76	174	3.8	2232	5	P91365
77	173.5	3.8	870	5	Q22822
78	173.5	3.8	2083	5	Q9N435
79	173.5	3.8	895	10	Q9N3H4
80	173	3.8	695	10	Q9S039
81	173	3.8	353	5	Q8T776
82	172.5	3.8	1233	11	P97820
83	172	3.8	664	4	Q14515
84	172	3.8	1007	4	Q43379
85	172	3.8	1157	5	Q9VTA7
86	171.5	3.8	1964	2	Q59947
87	171.5	3.8	4498	5	Q9W223
88	171	3.8	819	5	Q8SXM2
89	171	3.8			

Q8SY55 drosophila  
Q9HBE2 homo sapien  
Q96JK2 homo sapien  
Q9V517 drosophila  
Q95TD6 drosophila  
Q9B115 drosophila  
Q9FWE0  
Q60336 homo sapien  
Q17490 caenorhabdi  
Q9VTK8 drosophila  
Q960U6 drosophila  
Q9V054 drosophila  
Q17343 caenorhabdi  
Q23330 caenorhabdi  
Q23587 caenorhabdi  
Q9PGL1 fugu rubrip  
Q9VB40 drosophila  
Q9VAV4 drosophila  
Q9UAO2 caenorhabdi  
Q9P785 schizosacch  
Q9S255 arbidopsis  
Q67643 gallid hep  
Q9W320 drosophila  
Q9R010 mus musculu  
Q96A69 homo sapien  
Q96ST2 homo sapien  
Q8RCV4 homo sapien  
Q8TRT0 homo sapien  
Q9G0R0 drosophila  
Q970P7 streptococ  
Q8V099 staphylococ  
Q95819 homo sapien  
Q8WSM1 caenorhabdi  
Q8R1B6 dictyosteli  
Q86167 salmonele  
Q9VPL9 drosophila  
Q9N164 drosophila  
Q18918 caenorhabdi  
Q990Y4 streptococ  
Q9HGK6 candida alb  
Q75172 homo sapien  
Q9FNDS arbidopsis  
Q9NST7 homo sapien  
Q9LXN4 arbidopsis  
Q9CH86 lactococcus  
Q57576 cynops pyr  
Q19764 caenorhabdi  
Q76891 caenorhabdi  
Q90YE6 xenopus lae  
Q17084 caenorhabdi  
Q9DC88 morone saxa  
Q9VC00 drosophila  
Q9N541 caenorhabdi  
Q94317 schizosacch  
Q21418 schizosacch  
Q22248 caenorhabdi  
Q8T1A0 dictyosteli  
Q9C8N3 arbidopsis  
Q25388 loligo peai  
P91365 caenorhabdi  
Q22822 caenorhabdi  
Q9N435 mus musculu  
Q9N3H4 caenorhabdi  
Q9S039 arbidopsis  
Q8T776 brachiosteo  
P97820 mus musculu  
Q14515 homo sapien  
Q43379 homo sapien  
Q9VTA7 drosophila  
Q59947 streptococ  
Q9W223 drosophila  
Q8SXM2 drosophila

90	170.5	3.8	679	5	095XW8	095XW8 caenorhabdit
91	170.5	3.8	1010	5	09YB01	09YB01 drosophila
92	170.5	3.8	3198	5	09YB08	09YB08 manduca sex
93	170.5	3.8	2099	4	09YB08	09YB08 homo sapien
94	170.5	3.8	2099	4	09Y2W9	09Y2W9 homo sapien
95	169.5	3.7	619	13	09Y198	09Y198 oncorhynch
96	169.5	3.7	625	2	09Y198	09Y198 staphylococ
97	169.5	3.7	1763	11	09YJX5	09YJX5 mus musculu
98	169	3.7	844	2	09R235	09R235 shigella so
99	169	3.7	1255	2	09R236	09R236 shigella so
100	169	3.7	1421	5	09Y644	09Y644 plasmodium
101	169	3.7	3257	5	09Y736	09Y736 drosophila
102	168.5	3.7	576	2	09YKJ3	09YKJ3 enterococcu
103	168.5	3.7	664	4	014800	014800 homo sapien
104	168.5	3.7	961	3	09Y223	09Y223 emericeila
105	168.5	3.7	1218	5	09Y6H7	09Y6H7 drosophila
106	167.5	3.7	2342	5	046048	046048 drosophila
107	167	3.7	1402	5	095VT1	095VT1 plasmodium
108	167	3.7	3571	10	09SL27	09SL27 arabidopsis
109	167	3.7	3574	10	09AUB4	09AUB4 arabidopsis
110	166.5	3.7	862	11	09YK1	09YK1 rattus norv
111	166.5	3.7	1693	5	09W3P3	09W3P3 drosophila
112	166.5	3.7	2722	5	019135	019135 caenorhabdi
113	166	3.7	643	5	017907	017907 caenorhabdi
114	166	3.7	763	5	0960Y7	0960Y7 drosophila
115	165.5	3.7	891	10	09ZWO8	09ZWO8 arabidopsis
116	165.5	3.7	894	10	09FYB2	09FYB2 arabidopsis
117	165.5	3.7	1475	5	025842	025842 plasmodium
118	165	3.6	417	4	08WXX6	08WXX6 plasmodium
119	165	3.6	721	2	09F866	09F866 enterococcu
120	165	3.6	970	11	08VBY1	08VBY1 rattus norv

## ALIGNMENTS

RESULT 1

09DC22 PRELIMINARY: PRT: 876 AA.

AC 09DC22: 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE 1200060581K protein.

GN 1200060581K.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI\_TaxID=10090;

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=LUNG;

RX MEDLINE=1085660; PubMed=11217851;

RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Fukuda S., Akawa J., Hara A., Fukunishi Y., Komono H., Adachi I., Fukuoka K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K., Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R., Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fleischmann W., Gaasterland T., Glaser C., King B., Kochava H., Kuehl P., Lewis S., Matsuo Y., Nakado I., Pesole G., Quackenbush J., Schram L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G., Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P., Mordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F., Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L., Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kontsuki S., Hayashizaki Y.

RT Functional annotation of a full-length mouse cDNA collection.;

RL Mature 409:65-690(2001).

CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).

DR EMBL: AK004618; BAB23414.1; -.

DR MGD: MGI:1921356; 1200060581K.

DR InterPro: IPR000048; IO-region.

DR InterPro: IPR001005; MyD\_DNA\_binding.

DR InterPro: IPR001680; WD40.

DR Pfam: PF00612; IQ; 1.

DR Pfam: PF00400; WD40; 7.

DR SMART: SM00015; IQ; 1.

DR SMART: SM00320; WD40; 7.

DR PROSITE: PS00037; MYB\_1; UNKNOWN\_1.

DR PROSITE: PS02944; WD\_REPEATS\_REGION; 2.

KW Repeat; WD repeat.

SQ SEQUENCE 876 AA; 97587 MW; 677B1A3F8DFB407 CRC64;

Query Match 87.7%; Score 3972; DB 11; Length 876;

Best Local Similarity 86.5%; Pred. No. 3; I.e-250;

Matches 764; Conservative 27; Mismatches 62; Indels 30; Gaps 4;

QY 1 MSRGSGYPLMLMDVRRKSLGLEDDPSRLRSRYLGRREFIQRKLEATLNVHDCGVNTICWN 60

DB 1 MARSGCPHLMVDRKRSGLLEDDPSRLRSRYLGRREFIQRKLEATLNVHDCGVNTICWN 60

QY 61 DREGYLLSGDDPKLYISNYSRKYLTTRSGRANFSAKFLPCINDQIVSCSGDGI 120

DB 61 DREGYLLSGDDPKLYISNYSRKYLTTRSGRANFSAKFLPCINDQIVSCSGDGI 120

QY 121 FYINVEDAETNQCOCFTCHYTTEIMVNDPYFLSCGEDGYRWFDTRIKTSCTKE 180

DB 121 FYINVEDAETNQCOCFTCHYTTEIMVNDPYFLSCGEDGYRWFDTRIKTSCTKE 180

QY 121 FYINVEDAETNQCOCFTCHYTTEIMVNDPYFLSCGEDGYRWFDTRIKTSCTKE 180

DB 121 FYINVEDAETNQCOCFTCHYTTEIMVNDPYFLSCGEDGYRWFDTRIKTSCTKE 180

QY 181 DCKDDILINCRRAATVAICPPPIYTLAVCCSSSVRIYRRMLGRATGNTVAGRGITGM 240

DB 181 DCKDDILINCRRAATVAICPPPIYTLAVCCSSSVRIYRRMLGRATGNTVAGRGITGM 240

QY 241 VAFPIPSHLNKKSCRYTSLCYSEDEGEILYSYSDYTYLPDPKDDTARELKTSAEERE 300

DB 241 VAFPIPSHLNKKSCRYTSLCYSEDEGEILYSYSDYTYLPDPKDDTARELKTSAEERE 300

QY 301 ELRQPPYKRLRLGDWSDTGPRAPESEERDEQSPNVSLMORMSDMLSRWEEASEVA 360

DB 301 ELRQPPYKRLRLGDWSDTGPRAPESEERDEQSPNVSLMORMSDMLSRWEEASEVA 360

QY 361 OSNRGRGRPRGTGSDSTLTPTVSPDLAESEFAMVDPAEQFLPSTSTSAQ 420

DB 361 OSNRGRGRPRGTGSDSTLTPTVSPDLAESEFAMVDPAEQFLPSTSTSAQ 420

QY 421 AHSTRSPESPHTPLISSPDSEQRQVEASGHTTHQSDNNNEKLSPKRGTPVLSLH 480

DB 421 AHSTRSPESPHTPLISSPDSEQRQVEASGHTTHQSDNNNEKLSPKRGTPVLSLH 480

QY 481 YSTGTTSTIKLFTDEMSSIASSSRGISGHCKSEGOEFTVQSSVQPPREGSETRAP 540

DB 481 YSTGTTSTIKLFTDEMSSIASSSRGISGHCKSEGOEFTVQSSVQPPREGSETRAP 540

QY 541 EESSEDTYKQEGYASNPENHINITQSDFKTAPLDSNGERNDLNDRSCVPEESA 600

DB 541 EESSEDTYKQEGYASNPENHINITQSDFKTAPLDSNGERNDLNDRSCVPEESA 600

QY 540 EELSE-----KGLPENLQNOIDTQADLDFAPLPSDNGEANNNSQSDPGLPREGT 593

DB 540 EELSE-----KGLPENLQNOIDTQADLDFAPLPSDNGEANNNSQSDPGLPREGT 593

QY 601 SSEKAKE-----PETSQDT-----STESATVNNNTNPPQQTETATG 637

DB 601 SSEKAKE-----PETSQDT-----STESATVNNNTNPPQQTETATG 637

QY 594 LSEFDRECEQASTESATRNASTRKPELPSQTEALQASTESATRNASTRKPELPSQTEAL 653

DB 594 LSEFDRECEQASTESATRNASTRKPELPSQTEALQASTESATRNASTRKPELPSQTEAL 653

QY 638 PSAHEESTRPSALQDDDDDDDDVLLPGARYRAGPGGRSVAVARIGOFFRRRERKEME 697

DB 638 PSAHEESTRPSALQDDDDDDDDVLLPGARYRAGPGGRSVAVARIGOFFRRRERKEME 697

QY 654 PLAEHDSARSAALQDDDDDDVLLPGARYRAGPGGRSVAVARIGOFFRRRERKEME 713

DB 654 PLAEHDSARSAALQDDDDDDVLLPGARYRAGPGGRSVAVARIGOFFRRRERKEME 713

QY 698 ELDTLNTLRPLVKKVYVYGHRSRTMIKEANFWAGNFWAGNFWAGNFWAGNFWAGNFWAGN 757

DB 698 ELDTLNTLRPLVKKVYVYGHRSRTMIKEANFWAGNFWAGNFWAGNFWAGNFWAGNFWAGN 757

QY 714 ELDTLNTLRPLVKKVYVYGHRSRTMIKEANFWAGNFWAGNFWAGNFWAGNFWAGNFWAGN 773

DB 714 ELDTLNTLRPLVKKVYVYGHRSRTMIKEANFWAGNFWAGNFWAGNFWAGNFWAGNFWAGN 773

QY 758 EADNHVVCLOPHPPDPLASGSDYDIKTSPLSESRFNKRLADEYITNEMLEETR 817

DB 758 EADNHVVCLOPHPPDPLASGSDYDIKTSPLSESRFNKRLADEYITNEMLEETR 817



Db 774 EADNHVNCVLPHPDPLIASSGIDYDIKIMSPLEESIRFKRLADEVITNMLLESTR 833  
 QY 818 NTIVPASFMRLMLASLNHTRADLRGDRSGSGQENENDEE 860  
 Db 834 NTIVPASFMRLMLASLNHTRADLRGDRSGSGQENENDEE 876

## RESULT 2

08TB19  
 ID 08TB19 PRELIMINARY; PRT; 513 AA.

AC 08TB19; 01-JUN-2002 (Tremblrel. 21, Created)  
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Similar to PC326 protein.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NC NCBL\_Taxid-9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-PRIMARY B-CELLS;  
 RA Strausberg R.;  
 RL Submitted (Mar-2002) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; BC025262; AAH25262.1;  
 SQ SEQUENCE 513 AA; 56814 MW; 0D00CB9319C00F8 CRC64;

Query Match 58.8%; Score 2663; DB 4; Length 513;  
 Best Local Similarity 99.8%; Pred. No. 2.5e-165;  
 Matches 512; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 348 MLSRFEAESEVAQSNRGRSRPGTSGDSISITLPTVPSSPDLEVSTAMEVDTPAQ 407  
 Db 1 MLSRFEAESEVAQSNRGRSRPGTSGDSISITLPTVPSSPDLEVSTAMEVDTPAQ 60  
 QY 408 FLQPTSTSTMSQAQAHSTSPSPSTPLSSPDSEQROSVASGHHTHQSDNNKLS 467  
 Db 61 FLQPTSTSTMSQAQAHSTSPSPSTPLSSPDSEQROSVASGHHTHQSDNNKLS 120  
 QY 468 PKPGTGEVPLSLHYSTEGTSTSTIKLNTDEMSSIASSRGISGCKSGQESFVPOS 527  
 Db 121 PKPGTGEVPLSLHYSTEGTSTSTIKLNTDEMSSIASSRGISGCKSGQESFVPOS 180  
 QY 528 VQPPGDESETAPRESSSDVTKYQGVSAENPVENHINITOSDKFTAKPLDNGSERNDL 587  
 Db 181 VQPPGDESETAPRESSSDVTKYQGVSAENPVENHINITOSDKFTAKPLDNGSERNDL 240  
 QY 588 NLDSSCGVPEESASEKKEPETSDQSTESATNENNTNPEPQFTEATGSAHEETSTR 647  
 Db 241 NLDSSCGVPEESASEKKEPETSDQSTESATNENNTNPEPQFTEATGSAHEETSTR 300  
 QY 648 DSALQDITDDDDDDVLLPGARYRAGPGDRSAVARIOEFFRRKRKMEELDTNTRP 707  
 Db 301 DSALQDITDDDDDDVLLPGARYRAGPGDRSAVARIOEFFRRKRKMEELDTNTRP 360  
 QY 708 LVKAVYVGHNSRRTMKEANFWGANFWASGDCGHIFIMDRHTAEHLMLLEADNHVNCV 767  
 Db 361 LVKAVYVGHNSRRTMKEANFWGANFWASGDCGHIFIMDRHTAEHLMLLEADNHVNCV 420  
 QY 768 QPHFDPILASSGIDYDIKIMSPLEESIRFKRLADEVITNMLLESTRNTIVPASFM 827  
 Db 421 QPHFDPILASSGIDYDIKIMSPLEESIRFKRLADEVITNMLLESTRNTIVPASFM 480  
 QY 828 LKMLASLNHTRADLRGDRSGSGQENENDEE 860  
 Db 481 LKMLASLNHTRADLRGDRSGSGQENENDEE 513

RESULT 3  
 ID 09P000 PRELIMINARY; PRT; 533 AA.  
 AC 09P000; 09HOM1;  
 DT 01-OCT-2000 (Tremblrel. 15, Created)

DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE PC326 protein.  
 GN PC326 OR DKFZP434C2211.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NC NCBL\_Taxid-9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-ADRENAL GLAND;  
 RA Li Y., Shi J., Huang C., Jiang C., Ren S., Zhou J., Yu Y., Xu S.,  
 Wang Y., Fu G., Chen Z., Han Z.,  
 RT "A novel gene expressed in human adrenal gland."  
 RL Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 52-533 FROM N.A.  
 RC TISSUE-TESTIS;  
 RX MEDLINE-21154917; PubMed-11230166;  
 RA Mennig S., Well B., Wellenreuther R., Gassenhuber J., Glassl S.,  
 Mennig S., Boecker M., Bloeker H., Bauersachs S., Blum H.,  
 Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,  
 Mewes H.W., Ottenwälder B., Obermaier B., Tampe J., Heubner D.,  
 Wambolt R., Korn B., Klein M., Poustka A.;  
 RT "Towards a Catalog of Human Genes and Proteins: Sequencing and  
 RT Analysis of 500 Novel Complete Protein Coding Human cDNAs."  
 DR EMBL; A1150234; AF077471.1;  
 DR EMBL; A1136738; CAB66672.1;  
 DR InterPro: IPR000048; IO\_region.  
 DR InterPro: IPR001005; Myb\_DNA\_binding.  
 DR Pfam: PF00612; IO: 1.  
 DR Pfam: PF00400; WD40. 2.  
 DR SMART: SM00015; IO: 1.  
 DR SMART: SM00320; WD40. 2.  
 DR PROSITE: PS00037; MYB\_1; UNKNOWN\_1.  
 DR PROSITE: PS02094; WD\_REPEATS\_REGION; 1.  
 KW Hypothetical protein; Repeat; WD repeat.  
 SQ SEQUENCE 533 AA; 58974 MW; 7A3816892447A77F CRC64;

Query Match 58.5%; Score 2647; DB 4; Length 533;  
 Best Local Similarity 96.2%; Pred. No. 2.9e-164;  
 Matches 513; Conservative 0; Mismatches 0; Indels 20; Gaps 1;

QY 348 MLSRFEAESEVAQSNRGRSRPGTSGDSISITLPTVPSSPDLEVSTAMEVDTPAQ 407  
 Db 1 MLSRFEAESEVAQSNRGRSRPGTSGDSISITLPTVPSSPDLEVSTAMEVDTPAQ 60  
 QY 408 FLQPTSTSTMSQAQAHSTSPSPSTPLSSPDSEQROSVASGHHTHQSDNNKLS 467  
 Db 61 FLQPTSTSTMSQAQAHSTSPSPSTPLSSPDSEQROSVASGHHTHQSDNNKLS 120  
 QY 468 DSALQDITDDDDDDVLLPGARYRAGPGDRSAVARIOEFFRRKRKMEELDTNTRP 707  
 Db 301 DSALQDITDDDDDDVLLPGARYRAGPGDRSAVARIOEFFRRKRKMEELDTNTRP 360  
 QY 708 LVKAVYVGHNSRRTMKEANFWGANFWASGDCGHIFIMDRHTAEHLMLLEADNHVNCV 767  
 Db 361 LVKAVYVGHNSRRTMKEANFWGANFWASGDCGHIFIMDRHTAEHLMLLEADNHVNCV 420  
 QY 768 QPHFDPILASSGIDYDIKIMSPLEESIRFKRLADEVITNMLLESTRNTIVPASFM 827  
 Db 421 QPHFDPILASSGIDYDIKIMSPLEESIRFKRLADEVITNMLLESTRNTIVPASFM 480  
 QY 828 LKMLASLNHTRADLRGDRSGSGQENENDEE 860  
 Db 481 LKMLASLNHTRADLRGDRSGSGQENENDEE 513

OY 748 RHTAEHMLLEADNHVNCIOPHPDPIIASGIDYDIKIMSPLEERSIFNRKLADEVIT 807  
 DB 421 RHTAEHMLLEADNHVNCIOPHPDPIIASGIDYDIKIMSPLEERSIFNRKLADEVIT 480  
 OY 808 RNEIMLEETRNITVPAFSLMLRLASLNIHRAIRLGRSGSGGONENDEE 860  
 DB 481 RNEIMLEETRNITVPAFSLMLRLASLNIHRAIRLGRSGSGGONENDEE 533

RESULT 4  
 ID Q90UD5 PRELIMINARY; PRT; 485 AA.

AC Q90UD5; 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE D0703H14.1 protein (Fragment).  
 GN D0703H14.1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RA Graftam D.;  
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).  
 DR EMBL; AL031287; CAB52430.1;  
 DR InterPro; IPR001680; WD40.  
 DR Pfam; PF00400; WD40; 5.  
 DR SMART; SM00320; WD40; 5.  
 DR PROSITE; PS50294; WD\_REPEATS\_REGION; 1.  
 DR Repeat; WD repeat.  
 KM NON\_TER  
 FT SEQUENCE 485 AA; 54745 MM; F47AF430351AE219 CRC64;

Query Match  
 Best Local Similarity 49.9%; Score 2258; DB 4; Length 485;  
 Matches 426; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 34 RREFIORLKLKLEATLVHDCVNTICMNDTGEYILSGSDTKIVINPYSRKYLTTIRSGH 93  
 DB 1 RREFIORLKLKLEATLVHDCVNTICMNDTGEYILSGSDTKIVINPYSRKYLTTIRSGH 60  
 OY 94 RANIFSAKFLPCTNDKQIVSCSGDGYIFTYNEQDAETNRQCOFTCHYTYEIMTVPD 153  
 DB 61 RANIFSAKFLPCTNDKQIVSCSGDGYIFTYNEQDAETNRQCOFTCHYTYEIMTVPD 120  
 OY 154 PYFPLSCGEGDGYRWFEDTRIKTSCTKEDCKDILLNCRRAATVAICPPITYLAVGSD 213  
 DB 121 PYFPLSCGEGDGYRWFEDTRIKTSCTKEDCKDILLNCRRAATVAICPPITYLAVGSD 180  
 OY 214 SSYRIYDRRLGTRATGNVAGRTGTGMVAFIPSHLNNSCWTSLCYEEDGOELLVYS 273  
 DB 181 SSYRIYDRRLGTRATGNVAGRTGTGMVAFIPSHLNNSCWTSLCYEEDGOELLVYS 240  
 OY 274 SDYIYLPDPTARELKTSAERERELROPVYKRLRLGDMSDTGPAPARESERERDG 333  
 DB 241 SDYIYLPDPTARELKTSAERERELROPVYKRLRLGDMSDTGPAPARESERERDG 300  
 OY 334 EGSBNVSLMORMSDMLSRWFEASVYAQSNRGRGRPRGCTGOSDITSLPYPSPDLE 393  
 DB 301 EGSBNVSLMORMSDMLSRWFEASVYAQSNRGRGRPRGCTGOSDITSLPYPSPDLE 360  
 OY 394 VSEIYAVDTTPAQFLQPSSTMSAQAHSPTSPHSPTPLSPDSBOROVASGH 453  
 DB 361 VSEIYAVDTTPAQFLQPSSTMSAQAHSPTSPHSPTPLSPDSBOROVASGH 420  
 OY 454 HTHHOSD 460  
 DB 421 HTHHOSD 427

RESULT 5  
 ID Q9LEU6 PRELIMINARY; PRT; 754 AA.

AC Q9LEU6; 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Hypothetical 83.5 kDa protein.  
 GN T30N20.210.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OC NCBI\_TaxID=3702;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RA Bevan M., Peters S.A., van Staveren M., Dirkse W., Stiekema W.,  
 RA Bancroft I., Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.;  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP DOMAINS).  
 DR EMBL; AL365234; CAB96849.1;  
 DR InterPro; IPR00345; CytC\_heme\_bind.  
 DR InterPro; IPR001005; Myb\_DNA\_binding.  
 DR InterPro; IPR001680; WD40.  
 DR Pfam; PF00400; WD40; 6.  
 DR PRODOM; PD000018; WD40; 1.  
 DR SMART; SM00320; WD40; 6.  
 DR PROSITE; PS00190; CYTOCHROME\_C; UNKNOWN\_1.  
 DR PROSITE; PS00037; MYB\_1; UNKNOWN\_1.  
 DR PROSITE; PS50082; WD\_REPEATS\_2; 2.  
 DR PROSITE; PS50294; WD\_REPEATS\_REGION; 2.  
 KM Hypothetical protein; Repeat; WD repeat.  
 FT SEQUENCE 754 AA; 83526 MM; C63FC6E7608B2101 CRC64;

Query Match  
 Best Local Similarity 13.0%; Score 590.5; DB 10; Length 754;  
 Matches 207; Conservative 11; Mismatches 279; Indels 245; Gaps 28;

OY 22 EDPFRLRSRYLGRREFIORLKLKLEATLVHDCVNTICMNDTGEYILSGSDTKIVINPY 80  
 DB 20 QDPHEVDQRKQFISLVRSLSQEELGEGHGCNMLAMNSGSLISGSDDLRINIMNY 79  
 OY 81 YSRKYLTTIRSGHRANIPSAFLPCTNDKQIVSCSGGVY--FTTN-----VEQDAETNR 133  
 DB 80 SSRKLHSDIDGHTANIFCTFVPEPESDDELVSAGAEVRLFTSLSGRAEDDNAIIP 139  
 OY 134 QCOFTCHYTYEIMTVPDNDYFTLSCGEGDGYRWFEDTRIKTSCTKEDCKDILLN 188  
 DB 140 SALVQCTHRRYKKAIVAGBNVYWSASDEGTILQOHFRETSCPPAGTIAHQEGR-SVLL 198  
 OY 189 NCRRAATVAICPP-----IPYLLAVGSGDSSVRIYDRMLGTRATG----- 230  
 DB 199 DLRSAGKALADPPKQTLSTKSCDISATRPILLVGGSDAFARLYDRMPLPLSSKRKM 258  
 OY 231 -----NAGSGTGMVAFIPSHLNK-----SCVTSICYEEDGOELLVYSYIYFD 281  
 DB 259 PPPPCVNY-----FCPMLSERGRITLHTHTFSPNGEVLVLSGGEHYILMN 307  
 OY 282 PDDPTARELKTSAERERELROPVYKRLRLGDMSDTGPAPARESERERDEGOSPNVSL 341  
 DB 308 VNNGTGIMQYTPGDVNLFSF-----SNNLHDVESPQVVS- 342  
 OY 342 MORMSDMLSRWFEASVYAQSNRGRGRPRGCTGOSDITSLPYPSPDLE-EVSEIYAME 400  
 DB 343 -----TPQNGFHRSS--SNAATYKCKTELEIYLAQMSLE 373  
 OY 401 VDT-----PAQFLQPSSTMSAQAHSSTSP-----TESPSTPLSPDSE 443  
 DB 374 EGTDFYALAEANEVLAHSDIESALRHECLCTRAALLLRKKNDAH-----NAVROCH 429

QY 444 QRSVAVASCHHHHOSDNNNEKISPKRGCEPLSLHYSTEGTTSTIKINFTEWSSIA 503  
 Db 430 NARRIDASSFKAHYVSEALQOL-----GKCKEALDFATAQHMPSPADIAVAKESIK 483  
 QY 504 SSSRGISGSHCKSGEGESFVPOSSVOPPGDSSTK-----APESEEDVTYKQ 551  
 Db 484 RLQQAAG-----AKNET-----GAGTTRVLSLDILYRSEANSDDSHMSRSE 528  
 QY 552 -EGVSAENVENHININTOSDKFTAKPLDSNGERNDLNDRCGVPEESASSEKAKEPET 610  
 Db 529 RSDSDVEDELELDIOTLSDD-EGRDTDSN-MRSLNL-----RIHYGDDKPEE-NT 579  
 QY 611 SQTSTESATNNNTNPEQFOTATGSAHESTSRSDALDPTDSDDDPVLIPGARVY 670  
 Db 580 VMASSTGATSSSON-----DRTSYOPEGALD-----605  
 QY 671 AGPDGRSAVARIQEFRRKREKMEELDTLIRRPVLYKMYKGRNRTMIKEANFNG 730  
 Db 606 -----MKRRYVGHCVNGTIDIKQASFLG 627  
 QY 731 --ANFVMSGDCGHIFIMDRHTAEHLMLLEADNHVNCLOPPHPPILLASGIDYDIKIM 788  
 Db 628 QRGGEYIAGSDGGRWFIEKQGRIMKVLVGDSEVINCIOGHPFDSVATSGIDNTIKIM 687  
 QY 789 SP 790  
 Db 688 SP 689

RESULT 6  
 QY 094B03 PRELIMINARY; PRT: 757 AA.  
 AC 094B03:  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Hypothetical 83.8 kDa protein.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eustosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,  
 RA Dale J.M., Gibson H.A., Goldsmith A.D., Jiang P.X., Lee J.M.,  
 RA Chuach H.L., Tang C., Toriumi M., Yu G., Bowser L., Carninci P.,  
 RA Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T., Kamiya A.,  
 RA Karlin-Neumann G., Kawai J., Kim C., Koeseema E., Lam B., Lin J.,  
 RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,  
 RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Tracy S.E.,  
 RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.,  
 RT "Full length cDNA of gene T30N20.210/At5g10940 (GI:8979728).";  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 CC 1- SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP DOMAINS).  
 DR EMBL: AY039964; AAK64141.1;  
 DR InterPro: IPR000345; CytC\_heme\_bind.  
 DR InterPro: IPR001005; Myb\_DNA\_binding.  
 DR InterPro: IPR001680; WD40.  
 DR Pfam: PF00400; WD40; 6.  
 DR ProDom: PD000018; WD40; 1.  
 DR PROSITE: PS00190; CYTOCHROME\_C; UNKNOWN\_1.  
 DR PROSITE: PS00037; MYB\_1; UNKNOWN\_1.  
 DR PROSITE: PS50082; WD\_REPEATS\_2; 2.  
 DR PROSITE: PS50294; WD\_REPEATS\_REGION; 2.  
 KW Hypothetical protein; Repeat; WD repeat.  
 SO SEQUENCE 757 AA; 83830 MW; 12BE5AAC114056D CRC64;

Query Match 12.8%; Score 580; DB 10; Length 757;  
 Best Local Similarity 24.5%; Pred. No. 1.7e-29;  
 Matches 207; Conservative 111; Mismatches 279; Indels 248; Gaps 29;

QY 22 EDSRRLRSRYLGRREIORTKLEATLVNHDGCVNTICWMDTGTYILLSGDDTKLVISNP 80  
 Db 20 QDPSHEVDQMOHSHSLVRLSOEQLGQGVNMAVANSNSLLISGSDDERINIMNY 79  
 QY 81 YSRKVLTTISGHRANIFSAKFLPCTNDKQIVSCSGGVY--PYTN-----VQDAETNR 133  
 Db 80 SSRKLHSLIDTGHANIFCKFYPERISDELAVSGAGDAEYRLNTISLSRAEDNAIIP 139  
 QY 134 QCFPTHGCTTYIMVYNDPYTFLLSCGEDGVYRMDTRIKTSC-----TKEDCKDILI 188  
 Db 140 SALYQCHTRRVKRLAVEPNPNVWWSASEDGTIRHDFRESTCPAGTAHQCR--SVLL 198  
 QY 189 NCRRAATSVACPP-----IPYLAAGCSQSVRYRMRMIGTATG-----230  
 Db 199 DLRSAGKRALADPPKOTLSKCDISATRPHLLVGSDDFARLYRMLPIASSSRKM 258  
 QY 231 -----NYAGRTGGVAVARIPSHLNK--SCRYTSLCYSEDOELIVSYSSDYTYLFD 281  
 Db 259 PPPPCVNY-----FCPMHLSEGRNTLHLYHFSPMGDEVLLSYGHEHYTAN 307  
 QY 282 PKD---DTARELTPSAERRELRQPYKRLRLRDGWDSTGRAPRERERFDGQSPN 338  
 Db 308 VANGIGSTGIMOTTPEDVNLFSF-----SNLHDIYESPPQ 343  
 QY 339 VSLMRMSDMLSRWFEASEVAQSNRGRSRPRGTSQSDISTLPTVPSPL--EVSET 397  
 Db 344 VS-----TTPQNFHS--SNATVAKCTELVETAKW 373  
 QY 398 AMEVDI-----PAQTLQPTSSSTMSQAHSSTSP-----TESPSTPLSSP 440  
 Db 374 SLEEGTVFYALEANNEVLDAHSNDIESALRHCLOTFRALLLRKKNDAAH---MAVR 429  
 QY 441 DSEORVAVASGHHTFHOSDNNNEKISPKRGCEPLSLHYSTEGTTSTIKINFTEWSS 500  
 Db 430 DCHNARRIDASSFKAHYVSEALQOL-----GKCKEALDFATAQHMPSPADIAVAKY 483  
 QY 501 STASSRGISGSHCKSGEGESFVPOSSVOPPGDSSTK-----APESEEDVTYKQ 548  
 Db 484 SLKRLQAAG-----AKNET-----GAGTTRVLSLDILYRSEANSDDSHMS 528  
 QY 549 KYO-EGVSAENVENHININTOSDKFTAKPLDSNGERNDLNDRCGVPEESASSEKAKE 607  
 Db 529 RSDSDVEDELELDIOTLSDD-EGRDTDSN-MRSLNL-----RIHYGDDKPEE 580  
 QY 608 PETSQSTESATNNNTNPEQFOTATGSAHESTSRSDALDPTDSDDDPVLIPGARVY 667  
 Db 581 -NTVMASSTGATSSSON-----DRTSYOPEGALD-----608  
 QY 668 RYRAGPDGRSAVARIQEFRRKREKMEELDTLIRRPVLYKMYKGRNRTMIKEANFNG 727  
 Db 609 -----MKRRYVGHCVNGTIDIKQAS 627  
 QY 728 FWG--ANFVMSGDCGHIFIMDRHTAEHLMLLEADNHVNCLOPPHPPILLASGIDYDI 785  
 Db 628 FLGGGEYIAGSDGGRWFIEKQGRIMKVLVGDSEVINCIOGHPFDSVATSGIDNTIM 687  
 QY 786 KIWSP 790  
 Db 688 KIWSP 692

RESULT 7  
 QY 094527 PRELIMINARY; PRT: 809 AA.  
 AC 094527:  
 DT 01-MAY-1999 (TREMBLrel. 10, Created)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Putative guanine nucleotide binding protein beta subunit-like.  
 GN SPBC609.03.  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomyces.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972H-;  
 RA Seeger K., Harris D., Wood V., Rajandream M.A., Barrell B.G.;  
 RL Submitted (JAN-1999) to the EMBL/Genbank/DBP databases.  
 CC -1- SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).  
 DR EMBL: AL035226; CAA2832.1;  
 DR InterPro: IPR001230; Prenyl\_site.  
 DR InterPro: IPR001680; WD40.  
 DR Pfam: PF00400; WD40; 5.  
 DR Prodom: PD000018; WD40; 1.  
 DR SMART: SM00320; WD40; 5.  
 DR PROSITE: PS00294; PRENYLATION; UNKNOWN\_1.  
 DR PROSITE: PS00678; WD\_REPEATS\_1; UNKNOWN\_1.  
 DR PROSITE: PS50082; WD\_REPEATS\_2; 1.  
 DR PROSITE: PS50294; WD\_REPEATS\_REGION; 2.  
 KW Repeat: WD repeat.  
 SQ SEQUENCE 809 AA; 92498 MW; 4877443546BD057E CRC64;

Query Match 12.1%; Score 547.5; DB 3; Length 809;  
 Best Local Similarity 24.8%; Pred. No. 2.4e-27;  
 Matches 236; Conservative 124; Mismatches 308; Indels 285; Gaps 41;

27 LRSRYGRRERFOR-----LKLKATLVHDCGVTICWNTGTEYILSGSD 71  
 6 LLSRLQDYDMQROKTSRIQYNGSTWLTIDQLKELGHTGCVNLLMSADEPFLSGSD 65  
 72 DTKLV---ISNPYSKRYLTTRSGHRANITSAKFLPCTDKOIVSCSGDGVITYNVED 128  
 66 DTKLVYVDFNEKPRHL--ISTGHQVNIKSAFVYSNNRQILSASGDKLRLDSDS 123  
 129 -----AETNQCQPTCHGTYTILMVNPDPYFELSCGEGYRWDFRLKTSCTE 180  
 124 KEGMDHMETQTRC--WSCALDSKNIVPCDNK--HFLVCSBEGDTRKQYDIKPHVQND 181  
 181 -DCKDILLIN---CRRATSVAILCPPIPYLLAVGSCDSVRIRYRRLMTRA---TGNV 232  
 182 LDC-PSLIVNPNYRLNLTITWSPNPFYFALGTHPYAFIDRMVKKSPFADWTMT 240  
 233 AGRGTTGMVARTFPLSHLNKSC-----RYSLSCT--SEDOGLIVSYSSYIYFDK 283  
 241 SPKDCRCVAKFSP---DSCNSQGLIDRYITCCQFSANPPELLIVMSNDVYLFVHV 296  
 284 DDTARELKTPSAEERREELKOPVYKRLRLGWDSDTGPAPRAPEREREDGQSPNVSLMQ 343  
 297 EDKS---YTFTEKIEDSKNPKPSL-----LQVPLKRRK-----330  
 344 RMSDMSRWFEASEVAQSNRGRGRSRPRGGISQSDISTLPYPPSPDLVSETAMEVDT 403  
 331 ---YSPWY-----KNNFG-----AST 343  
 404 PARQFLQPSSTMSAQAHSSTPSPHSTPLSSPDSPQSVASGHHTHQSDNNK 463  
 344 PA-----SRYSNRPYTAAPRK-----HTYQAKENT 370  
 464 EKL--SPKPGTGEVPLSLHSTGTTSTIK--LNFTEMSIASSS-----506  
 371 EKFTTNGGLYESTIVSGRLSHFSRSIQYKDAIYFLENNTYIPDSNGLNHSIRVSALRY 430  
 507 -RGIGSHCKSEGESEFVPOSSVQPPRGDEFTAPRESSDVKYQEGV-----A 556  
 431 WRACVSLIALMDTVSLDEPTITIQAGWG---WLYDFNMWVTRYLGLISDHVALQNSPP 485  
 557 ENPVENHINTQSDKTFAPKPSNSGE-----RNDLN-----LDRSGVPE 597  
 486 TNVAKQNVLCDDPE-PSRVLFSNPSMIMAFARIDNDLSVRRFVHKVRLGCLLI 544  
 598 ESASSEKAKEP---ETSDQSTESATENN---TNEPOFQTEATGPS-----AH 641  
 545 SSDIYWEQFQWDSSTSDVTSISQRLDENGFLTLLEPPVNYEMNEVSSGENTIVSMYTG 604

642 EETSTRDSALQD-----TDSDDDPVLIPGARVRRAGDRRSARVARIQEEF 687  
 605 SLDLNDDDDDYQDESESYASDDEDDSEDDSEDEPTLLS-----642  
 688 RRRKRRKEMELDTNIRRLPYKMYKGRNSTRMTKMFAN--FVMSGSDCHIFT 745  
 643 LRMKRRKAVPNPVNTH---VKSXY-GHCNVSS-TKNVFNQNDYVMSGSDDRFT 697  
 746 WDRHTARHMLLEADNVVNCLOPHPPDPITLASSGIDYDIKWS-----PLESRITFNK 800  
 698 WDKLNASTILAIHGDSSAVAVIEHPHPCPLTAVSGIDSYIKIFINTENTPPSCGRNHTSN 757  
 801 LADEVITRNEMLTEPNTITVPASFMLRLASLNHTRADRLGDSGSGQ 853  
 758 -SYKIITMEKNRQGGSDSYI-TSRMLSHLAVRAH-----DDGCGHE 799  
 RESULT 8  
 ID Q9V8C7 PRELIMINARY; PRT; 628 AA.  
 AC Q9V8C7;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, last annotation update)  
 DE CG5124 protein.  
 GN CG5124.  
 OS Drosophila melanogaster (fruit fly).  
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;  
 CC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY.  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.A.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Butlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup I.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Dubin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,  
 RA Jaisli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodura C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattel B., McIntosh T.C., Meleod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclik J.M.,  
 RA Palazzolo M., Peltan G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svitzkas R., Tector C., Turner R., Venter L., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2185-2195(2000).

Query Match	11.2%;	Score 508.5;	DB 5;	Length 62b;
Best Local Similarity	23.2%;	Pred. No. 5.8e-25;		
Matches 191;	Conservative 107;	Mismatches 261;	Indels 265;	Gaps

[illegible]

```

Db 290 NMGYG----- 294
QY 423 SSSPESHSTPLSSPSEQRQSVASGHHTHHSDNNNEKLSPKPGTEPVLSLHYS 482
Db 295 ---SSPV----- 298
QY 483 TEGTSTIKLFTDMSIASSSRGSHCKSEGOESFVPOSSVQPPGDESEKPAEE 542
Db 299 ----- 298
QY 543 SSEDVTKYQGVSAENPVENHINTOSDKFYAKPLDSSNGERNDLNDSCGVPESSASS 602
Db 299 ----- 302
QY 603 EKAKEPESDQSTESATNENNTNPEPOTATGPAHEETSTDSALQDTDDSDDDV 662
Db 303 EKLQEME-----BPO----- 312
QY 663 LIPGARYRAGPGDRSARVARIOEFFRRRKERKEMEELDTLINRPLVKVYKGRNSRTM 722
Db 313 -----VYIGHRNAQT- 322
QY 723 IREANTGAGAN--FVMSGDCGHFTWDRHTAEHLMLLEADNHYVNCLOPHPEPDIASSG 780
Db 323 VGVNPFEGNDEYVTSQDGHIFIKKKGKLYRAMVGDRRVYNOLSESHPIPLASCG 382
QY 781 IDVDIKVSPLEESRIENKRIAEVITRNELMLEETRNITVPAFPM--LRMLASLHNR 838
Db 383 IESVAKLTPMSNDVLSLEPKIDKVELNRYGVEDOSRTVLTDPVIMHVLRLQROTSAF 442
QY 839 ADRLGDRSESGQENE-----NEDE 859
Db 443 TERRVYSTDIGSDGDFARISLAVANDE 472

RESULT 10
Q12839 PRELIMINARY; PRT; 597 AA.
AC Q12839;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, last annotation update)
DE H326 protein.
GN H326.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MYELOMA;
RA Bergsagel P.L.; Kuehl W.;
RT "H326 is a human gene homologous to murine PC326 that is ubiquitously
RT expressed."
RT submitted (FEB-1994) to the EMBL/GenBank/DBJ databases.
RL -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
CC EMBL; 006631; AAA16607.1; -
DR EMBL; IPR001680; WD40.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 7.
DR PRINTS; PR00320; GPROTEINBRPT.
DR PRODOM; PD000018; WD40; 1.
DR SMART; SM00320; WD40; 7.
DR PROSITE; PS50082; WD_REPEATS_2; 1.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW REPEAT; WD repeat.
SQ SEQUENCE 597 AA; 66384 MW; C0C696AD603645B9 CRC64;

```

Query Match 9.8%; Score 443.5; DB 4; Length 597;  
 Best Local Similarity 18.1%; Pred. No. 9.3e-21;  
 Matches 161; Conservative 83; Mismatches 173; Indels 471; Gaps 17;

```

QY 2 SRGSGYPHLLMD----VRKSLGLEDEPSRLRSYLGRRREFIQRILKEATLNVHDCVNTI 57
Db 142 SETSALPRPRWQALPALREBEL--SSARFVYACGARVYVHGRFHQHGEGHTGCVNTL 199
QY 58 CMTDGEYILISGSDDTIKLVISNPFYSKRVLTITSSGRANFSAKFLPCTDKQIVSCSD 117
Db 200 HFNQRTWLASGSDDLKVVYMDVWRQPVLDPESHGKSNVFOAKFLPNSGSDSTLANCARD 259
QY 118 GVIFYTNVEOD---AETNROCFTCHYGTVEIMTFVNDPYTFPLSGEDGTVMFPTR-- 172
Db 260 GQYRAVELSATQCCNKTKRAQ---HKGASHKALAPDSCTILSAGEDAVYFTIDLRD 316
QY 173 ---IKTSCYKEDCKDILINCRAPASVALCPPIPYLLAVGCSDSVRIYDRMLGTAT 229
Db 317 RPASKLVVTEKEK-----KVGLTYIVNPANTHOFAGVGRDQFRIYDQKIDENE- 368
QY 230 GNTAGKGTGMVARFIPSHLNKSCR--VTSLEYSGGQILVYSDDTYLLEDPKDDTA 287
Db 369 -----NNGVLKFCFPHHLYNSESANITICLVYSHDGETELASYNDEDIYLFN----- 415
QY 288 RELKTPSAEERBELRQPVKRLRLRGDMSDTGPARPESERERDQSPNVSLMQMSD 347
Db 416 ----- 415
QY 348 MLSRWEAESEVAQSNRGRSRPRRGTSQSDISTLPYVSPDLVEYSETAMEVDPAEQ 407
Db 416 ----- 415
QY 408 FLQPTSTSTMSAQAHSSTSPTEPHSTPLSSPDSERQSVASGHHTHHSDNNNEKLS 467
Db 416 ----- 415
QY 468 PKPGTEPVLSLHYSTEGTTSTIKLFTDMSIASSSRGSHCKSEGOESFVPOSS 527
Db 416 ----- 415
QY 528 VQPEGDSSTKAPDESEEDVTKYQEGVSAENPVENHINTIOSPKFYAKPLDSSNGERNL 587
Db 416 ----- 415
QY 588 NLDKSGVPEESASSEKAKPETSDQSTESATNENNTNPEPOTATGPAHEETSTR 647
Db 416 -----SSHSD----- 420
QY 648 DSALQDTDDSDDDPVILPGARYRAGPGDRSARVARIOEFFRRRKERKEMEELDTLINR 707
Db 421 -----GQY-----YKR- 427
QY 708 LVKMYTKGHRNSRTMIKEANFWG--ANFVNSGDCGHFTWDRHTAEHLMLLEAD-NHYV 764
Db 428 -----YKGHNNAT-VKGVNFYGPKESEFVVSQDCHIFLMEKSSCOIIOFMEDGKGAV 481
QY 765 NCLQPHPEPDIILASGDVYDIKIMSPLEESRIINRKLADEVITRNELMLEETRNITVPA 824
Db 482 NCLPPEPHPLVLTSGDHDVYKIMAPYAEASTELTGKVDIKKKNKRERDSDSLHQDTLFD 541
QY 825 SFMLMLASLHNRADR-----LEGDRSESGQENENDEE 860
Db 542 SHMLWFL--MHHLRQRHHHRMRPEVGATDADSDSESSDSDISEE 587

RESULT 11
Q01078 PRELIMINARY; PRT; 747 AA.
AC Q01078;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, last annotation update)
DE Protein PC326.
GN PET2 OR PROTEIN PC326.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

OX NCBI\_TaxID-10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BALB/C; TISSUE=PLASMACYTOMA;  
 RX MEDLINE=93026383; PubMed=1408147;  
 RA Eckhardt L.A., Laakov R., Kuehl W.M., Bergsagel P.L., Tindlin C.R.;  
 RT "Sequence in plasmacytomas of a murine cDNA encoding PC326, a novel gene  
 expressed in plasmacytomas but not normal plasma cells."  
 RL Oncogene 7:2059-2064(1992).  
 CC -1- SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP DOMAINS).  
 DR EMBL: M05564; AAA39895.1; -.  
 DR MGI:101758; Pet2.  
 DR InterPro: IPR001680; WD40.  
 DR Pfam: PF00400; WD40; 6.  
 DR ProDom: PD000018; WD40; 1.  
 DR SMART: SM00320; WD40; 5.  
 DR PROSITE: PS50082; WD\_REPEATS\_2; 1.  
 DR PROSITE: PS50294; WD\_REPEATS\_REGION; 1.  
 DR Repeat: WD repeat.  
 KW Sequence 747 AA; 83193 MW; 48E40B6378BD78A CRC64;

Query Match 9.5%; Score 429.5; DB 11; Length 747;  
 Best local Similarity 18.7%; Pred. No. 1e-19;  
 Matches 164; Conservative 76; Mismatches 173; Indels 465; Gaps 18;

OY 8 PHILMADV---RKRSLGLEDEPSRLSRYLGRREFIQRKLEATLVNHDGCVNTICMNDTG 63  
 DB PRPMNVLSALRDRLG--SSGRFYEACGALFVGRFLEHVEFGHSGCVNTHFNQHG 365  
 OY 64 EYLLSGSDTLYVSNPFSKRVLTIRSGHRANITSAPKLPCTNDKOIYSCSGDGYIFT 123  
 DB 366 TLLASGSDLVIVWDMKLSVLFNDSGKNNILQAKFLPCNDLILMCGRDGVRYA 425  
 OY 124 NVEDAETNROCOTCHYGYEIMTVNDPTFLSCGEGYVWMDTRIKTSCTEDCK 183  
 DB 426 QLSAVAGHMKRLVKHGGASHRLGLEPDPFRFLTSGDVAVENIDLQAHASK----- 481  
 OY 184 DDLI---NCRRAATVAICPPIDYILAVGCGSSSVRIYDRMIGTRATGNTAGCTTG 239  
 DB 482 --LVIKIDGDKRVGLYFVFNPNANYOFVAVGGDQFMRIYDQKIDENV-----NNG 531  
 OY 240 MVAFPIPSHL--NNKSCVTSICVSEDOEILVYSYSDIYLFDPKDTARELKTPSAE 297  
 DB 532 VLKFCFPHLLSSDPYPAHITSILMYSDTGELLASTNDEDIYF----- 574  
 OY 298 RREELQPPVRLRLRGDMWSDTGPRAPRERERDEQSPNVSLMQMSDMLSRWEAS 357  
 DB 575 ----- 574  
 OY 358 EVAQSNRGKSRPRGTSQSDISTLPYVSSPDLVSETAMEVDTPAEQFLQPSSTM 417  
 DB 575 ----- 574  
 OY 418 SAQAHTSSPSPHSTPLSSPDSBQSQVSEASGHHTHSDNNNEKLSPKPGTGPVL 477  
 DB 575 ----- 574  
 OY 478 SLVSTEGTTSTIKLNTDEMWSIASSSRGIGSHCKSEGGQESFVPOSSVOPPEGDSET 537  
 DB 575 ----- 574  
 OY 538 KAPESSSEDTYKQEGVSAENVEHNITQSDKFTAKPLDLSNGERNDLNDRCGVPE 597  
 DB 575 ----- 574  
 OY 598 ESASSEKAKEPETSDQSTESATNENNTNPEQFQTEATGSAHEETSTRDALSQDTDS 657  
 DB 575 -----NSDS 579  
 OY 658 DDDPVLIPGARYRAGGDRRSAAVARIQEFFRRRKRKMEELDTLNRPLVKNVYKGR 717  
 DB 580 D-----CAQY-----AKR-----YKGRH 592

OY 718 NSRTMKEANFWG--ANFVWSSDCHIFITMDRTAHEMLLEAD-NHVNCLQHPDP 774  
 DB 593 NNST-VKGVYFPRSEFVSGSGCHIFLWESSCOIVQPLEDEGCTINCISHPLP 651  
 OY 775 ILASGIDYDIKWSPLEESRIFNRKLA--DEVTRNLMEE--TRTITVPA--FM 827  
 DB 652 VLASGIDHEVKIWSPIAEF---SKLAGLKNVIKIKRKNDNTLRITSLFNNSMCL 708  
 OY 828 LRMILASLNH-----IRADRLGDRSGSGGGENENEDE 859  
 DB 709 MSHVQSGNYSRWSMRGIRINAGGDFSDSSSSSEETNOE 746

## RESULT 12

Q97001 PRELIMINARY; PRT; 493 AA.

ID Q97001  
 AC Q97001;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Hypothetical 54.9 kDa protein.  
 GN T12J5.10 OR AT4G35140.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bevan M., Murphy G., Ridley P., Hudson S., Hohenseel J., Mewes H.W.,  
 RA Mayer K.F.X., Schueller C.;  
 RA Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RA Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Murphy G., Ridley P., Hudson S., Mewes H.W., Lemcke K., Mayer K.F.X.;  
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Purnelle B., Boutry M., Goffeau A., Mewes H.W., Lemcke K.,  
 RA Mayer K.F.X.;  
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).  
 DR EMBL: AL035522; CAB36721.1; -.  
 DR EMBL: AL161586; CAB80231.1; -.  
 DR InterPro: IPR001680; WD40.  
 DR Pfam: PF00400; WD40; 6.  
 DR ProDom: PD000018; WD40; 1.  
 DR SMART: SM00320; WD40; 3.  
 DR PROSITE: PS50082; WD\_REPEATS\_2; 1.  
 DR PROSITE: PS50294; WD\_REPEATS\_REGION; 2.  
 KW Hypothetical protein; Repeat; WD repeat.  
 SQ Sequence 493 AA; 54937 MW; CF8F5C207CB48BAA CRC64;

Query Match 9.0%; Score 405.5; DB 10; Length 493;  
 Best local Similarity 18.9%; Pred. No. 2.1e-18;  
 Matches 162; Conservative 74; Mismatches 151; Indels 471; Gaps 20;

OY 9 HLMVRRKSLGLEDEPSRLSRYLGRREFIQRKLEATLVNHDGCVNTICMNDTGEYIIS 68  
 DB 17 HPYVDVFWREVGGISRRNFSASENVLRLRIYKLEKIKHKGCVNVSFNAEGDVLIS 76  
 OY 69 GSDTKLIVISNYSKRVLTITIRSGHRANITSAPKLPCTNDKOIYSCSGGCVIYFTVEOD 128  
 DB 77 GSDRVRVLMWQNLGNVLSFHSGHANNVFOAKFMFSDRIVTCAADGMF--DLRTE 133

```

QY 129 AETNRQCOPTCHYGTTEIMTVNPDPYTFELSCGEGTVAWFDTRIKTSCTKEDCKDDILLI 188
DB 134 APTC-----LFTC-----RSVPPRRRN-----MDAIOI 156
QY 189 NCRRAATSAICPPIPYLLAVGSCSDSVRIYDRMLGTATNGYAGRTGTVAA---RI 245
DB 157 N-----AIAIDPRSNLFAVGMGEYALYDIR-----RFOEGELNGFTTADHFC 202
QY 246 PSHL-NKSCRAVSLCYSDGGEILVSYSDYIYLPDPKDYARELKTPSAEERREELRQ 304
DB 203 PPHLIGNEVDVIGITGLAFSEQ-SELLVSYNDEFIYLFTP----- 239
QY 305 PPKVRLRLRGWSDTGPAPRAREPERERCEQSPNVSLMQRMSDMLSRWFEAEVAQSNR 364
DB 240 ----- 239
QY 365 GGRSRPRGATSGSDISTLPVFPSSPDLEVSFETAMEVDTPAQFLQPTSTSTMSQAHS 424
DB 240 GMGLG----- 244
QY 425 SSPTSPSTPLSSPDSEQRQSVASGHHHTHQSNNNEKLSPKPGTGEVPLSLHYSTE 484
DB 245 -----SNIPSPSPIS----- 254
QY 485 GTTSTIKLNTDEMSSIASSSRGIGSHCKSGQESFPQSSVQPPREGDSETKAPRESS 544
DB 255 -----KSPVSKS 261
QY 545 EDVTKYQGVSAENFVNHINITQDKFTAKPLDSNGERNDLNDRCGVPESASSEK 604
DB 262 -----ESSSPK 268
QY 605 AKPEPISQPTESATNENNTNPEQFOTATGSAHEETSTRSALODTDDSDDPVLI 664
DB 269 DE-----NHS----- 274
QY 665 PGARYRAGPDRRSARVARIQEFRRKRERKEMEELDTLINRRPLVKMYKGRNSRTMK 724
DB 275 -----VSLYKKGKNCET-VK 289
QY 725 EANTWG--ANTVMSGSCGHIIFINDRTAEHLMLIADNHNVCLOPHFPDPTIASSGID 782
DB 290 GVNFPGPSEVVGSDGRIPIWRKKGCELIRVMEADRHVNCIEPHPIVLAASGIE 349
QY 783 YDIKMSPLEBSRIENRKLADDEVITRNMLERNTTIVASFTKRLALSLN-HIRDR 841
DB 350 SDIKWT-----SKAERATLPE-----NIELPSRFRITPWLSELSFHDYDE 392
QY 842 LEGDRSGSGQENENEDE 859
DB 393 LFGNGMDIGIDNGEDE 410

```

## RESULT 13

```

Q9SZM8 PRELIMINARY; PRT; 426 AA.
AC Q9SZM8;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Hypothetical 47.5 kDa protein.
GN F20M13.40 OR A74G38480.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicot; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Medler H., Kutzner M., Wambutt R., Bancroft I., Mewes H.W.,
RA Mayer K.F.X., Scheller C.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
[2]

```

```

RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Medler H., Kutzner M., Wambutt R., Mewes H.W., Lemcke K.,
RA Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (WRP-ASP DOMAINS).
DR EMBL; AL035540; CAB37504.1; -.
DR EMBL; AL161593; CAB80512.1; -.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 7.
DR ProDom; PD000018; WD40; 1.
DR SMART; SM00320; WD40; 5.
DR PROSITE; PS50082; WD_REPEATS_2; 1.
DR PROSITE; PS50294; WD_REPEATS_REGION; 2.
KW Hypothetical protein; Repeat; WD repeat.
SQ
SEQUENCE 426 AA; 47451 MW; 0A6050076B8ECECA CRC64;

```

```

Query Match 8.5%; Score 384.5; DB 10; Length 426;
Best local similarity 17.6%; Pred. No. 4,1e-17;
Matches 137; Conservative 63; Mismatches 116; Indels 461; Gaps 14;

```

```

QY 32 LGRREFIQRKLEALLNHDGCVNTICMNDGIELLSGSDTKLYISNPKVLTTRS 91
DB 3 LOKQDLRLGLDKRLDKKCGVNTVSNADDDILLSSDDQVILIMQZASVKLSFDS 62
QY 92 GHRANIFSAKLPCTNDKQIVSCSGDGIYFYNEQDAETNQCQFTCHYGTTEIMTV 151
DB 63 GHENNIFOAKFMPFSDDBDTIYTSADKQYRSKILESQVETSL-LGKHQGVHKLAVEP 121
QY 152 NDPTFELSCGEGIVR-----WEDTRIKTSCTKEDCKDDILLINCRRA-----AT 195
DB 122 GSPSEFYTCGDEGAVKHCIFHFVFLSTRVATN-----LFTCKEAFNLVYLH 171
QY 196 SVAICPPIPYLLAVGSCSDSVRIYDRMLGTATNGYAGRTGTVAAFIISHLNKSCR 255
DB 172 AIAIDPRNPGILAAVAGMGEYARVDIR--STRSEDDHG----- 208
QY 256 VTSICYSDEGGEILVSYSDYIYLPDPKDYARELKTPSAEERREELRQPVKRLRLGD 315
DB 209 ITGLAFS-DQSELLASYSDEFIYLFTP----- 234
QY 316 WSDTGPAPRAREPERERCEQSPNVSLMQRMSDMLSRWFEAEVAQSNRGRSRPRGT 375
DB 235 -----DM-----GLG----- 239
QY 376 SQSDISTLPVFPSSPDLEVSFETAMEVDTPAQFLQPTSTSTMSQAHSPTSPHSTP 435
DB 240 -----PTPPSSSTK 248
QY 436 LLSSPDSEQRQSVASGHHHTHQSNNNEKLSPKPGTGEVPLSLHYSTEGTTSTIKLNF 495
DB 249 ----- 248
QY 496 TDEMSSIASSSRGIGSHCKSGQESFPQSSVQPPREGDSETKAPRESSDPTVYQEGVS 555
DB 249 -----TEERTPTP----- 256
QY 556 AENFVNHINITQDKFTAKPLDSNGERNDLNDRCGVPESASSEKAKEPEPISQTS 615
DB 257 ----- 256
QY 616 TESATNENNTNPEQFOTATGSAHEETSTRSALODTDDSDDPVLIIGARYRAGPD 675
DB 257 ----- 256
QY 676 RRSARVARIQEFRRKRERKEMEELDTLINRRPLVKMYKGRNSRTMKANFWG--ANF 733

```



Dd	257		-----	:   :       :	-YKHEHNRET -VKGVNFGPKCEY	279
Oy	734	VMSGSDCHFTWDRHTRFHLMLLEADNHVNCLOPHFPDPLIASSGDIDYDIKIWSP	790	:		
Dd	280	VVSQSGCRIFIRWKDKDELLRAMEADRHHVNVCIESHPHMPLMCSSGIDTIKIWTIP	336	: :              :   :           :   :               :   :		
<b>RESULT 14</b>						
OBSX92	ID	OBSX92	PRELIMINARY;	PRT:	743 AA.	
AC	08SX92;					
DT	01-JUN-2002	(TREMBLrel. 21, Created)				
DT	01-JUN-2002	(TREMBLrel. 21, Last sequence update)				
DE	01-JUN-2002	(TREMBLrel. 21, Last annotation update)				
CN	G28796p.					
CN	CG8001.					
OC	Drosophila melanogaster (Fruit fly).					
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;					
CC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;					
OX	NCHI_TaxID=7227;					
RN	[1]	SEQUENCE FROM N.A.				
RP	STRAIN=BERKELEY;					
RA	Stapleton M., Brokstein P., Hong L., Aghayani A., Carlson J.,					
RA	Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,					
RA	George R., Gonzalez M., Guarin H., Krommler B., Li P., Liao G.,					
RA	Miranda A., Mungall C.J., Nunoo J., Paclet J., Paragas V., Park S.,					
RA	Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rudin G.M.,					
RL	Celniker S.;					
DR	Submitted (Apr-2002) to the EMBL/GenBank/DDBJ databases.					
SQ	EMBL: AY094738; AACM1091.1; "					
	SEQUENCE 743 AA; 82751 MW; 96B713D71FC95EFF CRC64;					
Query Match                8.2%; Score 373; DB 5; Length 743;						
Best Local Similarity     17.3%; Pred.No. 4.8e-16;						
Matches 147; Conservative 73; Mismatches 163; Indels 468; Gaps						
Oy	31	YLGRREFIORLTLEALNVHDGCVAVTICWNDSGEYILSGSDDTKLVISNPYSRKVLTTIR	90	:   :       :		
Dd	301	YXSGLVFQEOMTLSSLINHGHCVSNLFNRAGDLICSGDDTLTYVMWAKEKOHRPR	360	: :              :   :           :   :		
Oy	91	SGHRANITSAKTLPCTNDKQLYSCGGDYIIETINEQDAETRKCQCFTHGYGTTEIMTV	150	: :              :   :           :   :		
Dd	361	SGHNNNIOTFKRIDASACLIDIVSSRDQGVRSAVLPSSGVATKLPIRLYTHSSSVHAIIV	420	: :              :   :           :   :		
Oy	151	PNDPPTYFLSCGGDGTRYMRVEDRIKTSCFEKDCKDILLNC-----RRATSVAICP	201	: :              :   :           :   :		
Dd	421	PLSRHELMSGADAAYKHFDLASNAAT-----TMRCYUNDENERGRVRLFSTAHPR	473	: :              :   :           :   :		
Oy	202	PIPYLIANGCSDSVARIYDRMKLTRATNGVARGTGTVAVARFIPSHLNKNSCVTSLSCY	261	: :              :   :           :   :		
Dd	474	YAPEPCVGS-SDDLILRVYDKRNL-AKAHQMAPRNLE-----LEAOITOLICAVY	520	: :              :   :           :   :		
Oy	262	SDDGOELIYVSDDYYLPDPKDDPARBELKTPSAEERRRELNQPPYKRLLRGDMWDGTP	321	: :              :   :           :   :		
Dd	521	NHSGETIIASTDAGITLYEDSKRYN-----RQDY-----	549	: :              :   :           :   :		
Oy	322	RAPESEREDEOSPNSVALRMQMDMLSFWEEASEVAOSNGRGRSPRGSTSODIS	381	: :              :   :           :   :		
Dd	550		549			
Oy	382	TLPYVSSPDLEVSETAMEVDPABEQILOPSTSTMSAQAHNTSSPTESHSTPLISSPD	441	: :              :   :           :   :		
Dd	550		549			
Oy	442	SEORQVEASGAHTTHHSDNNNEKLSPKPQTGPVSLHYSTEGTTTTIKLNFTDEMSS	501	: :              :   :           :   :		
Dd	550		549			
Oy	502	IASSSHIGISHCKSECOEESFYPOSSVOPEPGDSETKAPESESSEDYTKIOEGVSAENPVE	561	: :              :   :           :   :		

[illegible]

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodman T., Worley K.C., Wu D., Yang S., Yao O.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 CC -1 SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP DOMAINS).  
 DR EMBL: AE003472; AAF47563.1;  
 DR FlyBase: FB00035268; CG8001.  
 DR InterPro: IPR001680; WD40.  
 DR Pfam: PF00400; WD40; 6.  
 DR SMART: SM00320; WD40; 5.  
 DR PROSITE: PS50082; WD\_REPEATS\_2; 1.  
 DR PROSITE: PS50294; WD\_REPEATS\_REGION; 1.  
 KW Repeat: WD repeat.  
 SQ SEQUENCE 748 AA: 83334 MW: AC92C5F0E3ABFC1A CRC64;

Query Match 8.2%; Score 373; DB 5; Length 748;  
 Best Local Similarity 17.3%; Pred. No. 4,9e-16;  
 Matches 147; Conservative 73; Mismatches 163; Indels 468; Gaps 14;

QY 31 YLGRREFQRLKLEATLVNHDGCVNTICWDTGTYLISGSDTKLVISNPKVLTTR 90  
 DB 306 YGSRQVQVQWTLTSLVNHGCVNLSNFRAGDLCISGSDLTIVYDMKKEQLHFR 365  
 QY 91 SGHRANIPSAFLPCTNDKOIYSCSGDVIFITVNEQDAETNROQFCHGTTEIMTV 150  
 DB 366 SGHNNITQTFIDISAGLDIVSSRQGVRSVYPSGVIKPKRLTHSESHKILTV 425  
 QY 151 PNDPYTFLSGEDGTVMFDTIRIKTSCTKEDCKDILINC-----RRATSVATCP 201  
 DB 426 PSHRHEILMSAGEDAAVHKFDLRASNAAT-----TMRQVYNDENENGRVRLSIAHP 478  
 QY 202 PIPIYLVAGCSDSVRIYDRMLGTRATGNTAGKGTVMARFIPSHLNKSCRYTSLCY 261  
 DB 479 YAPFCVSG--SDILRYDKRNL-AKALHOMAPRNL-----LEAQITQITCAVY 525  
 QY 262 SEDQELIVSYSDYIYLFDPKDDTARELKTPSAEERREELRQPVKRLRSGMSDGP 321  
 DB 526 NHSSEILASVSDAGITLFDNRNN-----RGD----- 554  
 QY 322 RAPESERERDGEOSPUNVSLMQRMSDMLSRWEEASVVAQNRGRSRPGTGSODIS 381  
 DB 555 ----- 554  
 QY 382 TLPTVPSSDPLEVSETAMVDTPAFQFLQSTSTMSQAHSTSSPTSPHSTPLLSFD 441  
 DB 555 ----- 554  
 QY 442 SEQRQVSEAGHTHSDNNNEKLSPPKGTGEFVLSLHYSTEGTTSTIKLNTDEMSS 501  
 DB 555 ----- 554  
 QY 502 IASSRKGICGCKEGEGESFVPOSSVQPPGDESKAPESSSDVTIKYOGVSAENPVE 561  
 DB 555 -----LHC----- 557  
 QY 562 NHINIQSDKFTAKPLDSNGERNDLMDRSCGVPESSASERAKKEPETSQTSATN 621  
 DB 558 ----- 557  
 QY 622 ENNTNPEPQFQTEATGSAHEETSTRDSDLQDTPDDDDPVLIPGANYRAGPDRSVA 681  
 DB 558 ----- 557  
 QY 682 RIQEFFRRRERKEMBELDTLNIIRPLVKVYKGRHSRTMIKEANPWG--ANFVMSGD 739  
 DB 558 -----YKHINSRT-IGQVNFEGFRSEIIVSGSD 585

QY 740 CGHIFWDRHTAEHLMTLEADN-HVNCIQPHFPDPIILASSGDYDIKIMPLEESRIEN 798  
 DB 586 CGNIFPMQDNFALINYMGGDHVGNCLPEPMPVLTSLGHEHVKWTVOGPEPKLP 645  
 QY 799 RLADDEVITRN-----ELMLEETRNNTI-----TVPASFMRLMALNHRADRL 843  
 DB 646 EDILKQTLORNRNCNIVESIDDINNFGYIFRGLOGSASHLRHTSTSLGHQMDN-S 704  
 QY 844 GDRSGSGGEN 854  
 DB 705 SRSNGSNASN 715

RESULT 16  
 Q9NV87 PRELIMINARY; PRT; 615 AA.  
 AC Q9NV87.  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE CDNA FLJ10872 fis, clone NT2RP4001725, weakly similar to guanine  
 DE nucleotide-binding protein beta subunit.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Isegawa T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Ishibashi T., Fujimori K.,  
 RA Tanai H., Kimata M., Watanabe M., Hiraoaka S., Ishii S., Kawai Y.,  
 RA Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Negahari K.,  
 RA Masuho Y., Kanehori K.;  
 RT "NEDO human cDNA sequencing project."  
 RL Submitted (FE8-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1 SIMILARITY: CONTAINS 4 WD REPEATS (TRP-ASP DOMAINS).  
 DR EMBL: AK001734; BAA91868.1;  
 DR InterPro: IPR001405; MYB\_DNA\_binding.  
 DR InterPro: IPR001400; TPR.  
 DR InterPro: IPR001680; WD40.  
 DR Pfam: PF00515; TPR; 2.  
 DR Pfam: PF00400; WD40; 4.  
 DR Pfam: PD000018; WD40; 1.  
 DR ProDom: PD000018; WD40; 1.  
 DR SMART: SM00320; WD40; 3.  
 DR PROSITE: PS00037; MYB\_1; UNKNOWN\_1.  
 DR PROSITE: PS50082; WD\_REPEATS\_2; 2.  
 DR PROSITE: PS50294; WD\_REPEATS\_REGION; 1.  
 KW Repeat: WD repeat.  
 SQ SEQUENCE 615 AA: 69049 MW: D5930944AB691FF CRC64;

Query Match 7.3%; Score 330; DB 4; Length 615;  
 Best Local Similarity 24.3%; Pred. No. 2,4e-13;  
 Matches 144; Conservative 79; Mismatches 275; Indels 94; Gaps 21;

QY 13 DYKRSIGLEDPSSRLRSRYLGRREFQRLKLEATLVNHDGCVNTICWDTGTYLISGSD 72  
 DB 9 DLIRQIKERGALSFEFRHYHTDPFIRRLGLEALQHGSCGVNCLENNKEKGDLLASGSD 68  
 QY 73 TKLVISNPKVLTIRSGHRANIPSAFLPCTNDKOIYSCSGDVIFITVNEQDAETN 132  
 DB 69 QHTVMPDLHKKLLSMHTGHTANIFSVKFLPHAGDRLLTGADSKVHVDLTVKETIH 128  
 QY 133 ROCQFTCHYGTTVEIMVNPDPYFLSCGEDGVTRMFDTIRIKTSCTKEDCKD-IDLIN- 189  
 DB 129 ---MGDHTNNVKKRIATAPMPNPTFWASAEDGLRQYDLR-----ENSRHSEVLDLT 178  
 QY 190 ---CRR---AATSVATCPPIPIYLVAGCSDSVRIYDRML-----GTRATGNTAGRT---- 237  
 DB 179 EYCGQLVEAKCIIVNPDNCLAVGASGPFVRLDIDIMHNRKSMQSPSAGVHTFCDR 238  
 QY 238 ---TGAVARFIPSHL-----NKKSCRY---TSLCYSEDQGLIYSSDYLYLFD- 281  
 DB 239 QKPLPDGAOYVAGHLPVKLPDYNRLRVLVATYVTFSPNGTELLVMGGEQVYLDLT 298



QY 825 SFMLKMLASLHNRADRLGRCRSGSGOENENDEE 860  
 DB 316 DPLEVMLNMGY---RITGLSGAGASDDEDSSE 347

## RESULT 19

ID 096JK2 PRELIMINARY; PRT; 958 AA.  
 AC 096JK2;  
 DT 01-DEC-2001 (TREMblrel. 19, Created)  
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)  
 DE KIAA1824 protein (Fragment).  
 GN KIAA1824.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 ON NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN;  
 RX MEDLINE-21245130; PubMed-11347906;  
 RA Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XX.  
 RT The complete sequences of 100 new cDNA clones from brain which code  
 RT for large proteins in vitro.";  
 RL DNA Res. 8:85-95(2001).  
 CC EMBL; AB058727; BAB47453.1;  
 DR InterPro: IPR001680; WD40.  
 DR Pfam: PF00400; WD40; 5.  
 DR PROSITE: PSS0082; WD\_REPEATS\_2; 3.  
 DR PROSITE: PSS0294; WD\_REPEATS\_REGION; 1.  
 KW Repeat; WD repeat.  
 FT NON\_TER  
 SQ SEQUENCE 958 AA; 105178 MW; CF6799191C9F7B6 CRC64;

Query Match 5.18; Score 232.5; DB 4; Length 958;  
 Best Local Similarity 20.5%; Pred. No. 9.5e-07;  
 Matches 170; Conservative 112; Mismatches 343; Indels 203; Gaps 32;

QY 16 KRSLGLEPRLSRVYGR-----EFIRKLEATLVN-----HDCGVNITCW 59  
 DB 18 KRRAGLGSGMSRVGFLSQRGLHDDPLLTDPQRRRLRCGRNLTKKLLGHFGCVNAIEF 77  
 QY 60 -NDGTETLSGSDDTKLIVSNP---YSRVLTTIRSGHRANIF-----SAKFLPCTN 107  
 DB 78 SNNGGQWLVSQDDEFLVLMHMEQALHSRYKPIQLKGEHSHNIPCLAFNSGNTKVFSGN 137  
 QY 108 DKQIVSGSGGVLFYFNVEDQAEFNRCQCTCHGTYEIMTVNDYTYLSCGEDCTVR 167  
 DB 138 DEQVIL-----HDVESSEFLDYFAHEDAVYGLSVSPVNDNITASSDDGRVL 184  
 QY 168 WFDTRITKCTKEDKDDILINCRATSVAIQPIPYVYLVGCGSDSVRIYDR----- 222  
 DB 185 IWDIR-ESPGEPC-----LANPFAHSHVMPNVEPRLLATANSKSGVGLMDIRKQSS 239  
 QY 223 -----MGTATGTVACRGTTGM-VANFISHLNN-----KSCRV 257  
 DB 240 LIRYVGNLSLOSAMSVFENSQTOLLALRRRLPVLIDISRLPVFOQMGYFNSCTMK 299  
 QY 258 SLICYSDEGRIIVSYSDY-IYLFDP-KKDTAREL-KTPSAERERELRQPPVKRLRLRG 314  
 DB 300 SCFFAGDRDQYILSGSDFELVYMRILPADDEAGIGVNVGAFWVLGHSSTYNQVAFNP 359  
 QY 315 D-----WSDGPRAPRESEREREQESPNVSLQMRMSDMLRPFEEASEV 359  
 DB 360 HTYWCSSGVEKTIKIMS---PYKPOCTDDLDGRIEDRCILYHETYS----- 407  
 QY 360 AGSNRGGRGSRPPGGS-QSDITLPTVSSPDLVSEIAMEVDTPAEOFLQSTSTMS 418  
 DB 408 LVYNSGSLSHDYANOSVQEDPRMAAFDLSVLRREIEGMSDSDSD-----LSESTILQLH 463

QY 419 AQAHSTSPTESPHSTPILLSPDSQROSVASGHHTHOSDNNEKLSPKRGTEPYLS 478  
 DB 464 AGVSRSGTYTSESSASLPRPP-----PYDESADNANH-----LGP----- 501  
 QY 479 LHYSTEGTTSTIKINTDEMSS-----IASSRGIGSHCKSGOE 520  
 DB 502 LRVTTNTVASTPPPTCEDAASRQORLSALRRYQDKRLALNSRDEENVCVELDTPD 561  
 QY 521 SEVPOSSVQPEGDESETAPESSEDYTKQEGVSA---ENPVENHIIITQSKRTA--K 575  
 DB 562 LF-FRPRSPSEDESSSSSSSEDEELNERRASTWORAMRRQRTTRDKPSADIK 620  
 QY 576 PLDSNSGERN---DLNDRSCGVPEESASSEKAKEPETSDQTSATNENNTPEPOF 631  
 DB 621 PNTYIGEDNDYDPIKYND-----DLSSSPITSPRSTLT-----EI 658  
 QY 632 QTEATGPAHEETSTRD-----SALDDTDSDDPYLIPGA---RYRAG----- 672  
 DB 659 QPSRASPTSDIESEVERKLYKAYKWLRYSYISNNKDETSLVGGEADGRAGSHKNDP 718  
 QY 673 -PGDRRSVARIQEFERRRKRKEMELDTLIRRLPYKMYKGRNS 719  
 DB 719 APSSKKECLINAAQRNODLPPEGCSKDTKEETPRTPNSGPHERS 766

## RESULT 20

ID 09V5L7 PRELIMINARY; PRT; 747 AA.  
 AC 09V5L7;  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)  
 DE CG12892 protein.  
 GN CAP1-105 OR CG12892.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 ON NCBI\_TaxId=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY.  
 RX MEDLINE-20196006; PubMed-10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sultion G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer V.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt J., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Plannkoch C., Baldwin D.,  
 RA Bailey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottler P.,  
 RA Burris K.C., Busam D.A., Butler H., Caden E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup I.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fiesler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hosteln D., Houston K.A., Howland T.J., Wei M.-H., Ibegyan C.,  
 RA Jalali M., Kalush F., Kapran G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mlshina N.V., Mobery C., Morris J., Moshirei A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazon M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Massarman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 CC -1- SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).  
 DR EMBL: AE003829; AAF58788.1; -  
 DR Flybase: FBgn0033526; Cafl-105.  
 DR InterPro: IPR001632; GpoteIn\_B.  
 DR InterPro: IPR001680; WD40.  
 DR Pfam: PF00400; WD40; 5.  
 DR PRINTS: PR00319; GPROTEINB.  
 DR ProDom: PD000018; WD40; 1.  
 DR SMART: SM00320; WD40; 5.  
 DR PROSITE: PS00678; WD\_REPEATS\_1; 1.  
 DR PROSITE: PS00082; WD\_REPEATS\_2; 3.  
 DR PROSITE: PS50294; WD\_REPEATS\_REGION; 1.  
 KW Repeat; WD repeat.  
 SQ SEQUENCE 747 AA; 83394 MW; 0CF0BA42FCA2D89D CRC64;

## Query Match

Best Local Similarity 4.7%; Score 213.5; DB 5; Length 747;  
 Matches 167; Conservative 103; Mismatches 282; Indels 257; Gaps 40;

OY 8 PHILW-----DYKRSKSLGLEDPS--RLNSR-----YLGRREFIQRLKLE--A 45  
 DB 6 PEISWNRDPLVSLDIQNGILGRSPTICRLASGGSDAHILYVNRSDAEVDIELA 65  
 OY 46 TLNVHGCYNTIGMNDTGEYIISGSDTKIVT-----SNPYSRKVLTT 88  
 DB 66 DISRHQRAVNAVWSNGELASGDSESVYFIKQAKADHEVNIIVADGCGSEODKEWILT 125  
 OY 89 IR--SGHRANIFSAKFLPCGNDRKOIYSGSGDGVIFTYNEQ-----127  
 DB 126 LKYLKRHRREDIYDLNMAP--NSQFLVSGSVYDNTAMLMVDHSGSLAIIIDHKYVOGVAM 133  
 OY 128 -----DAETNR-----OCQFCHYGVTEYIMTVPDYTF--LS 159  
 DB 184 DPCNQIATMSTDQRKRIEDANTKRVILHRYSK-----VLPVKDHEMHGKSMR 232  
 OY 160 CGEDGVNRFDRIKTSCKRECK-----DIL--IN-----CRAATSAVACP 201  
 DB 233 LYQDGTILQFFRRL--CCTPDCKLLTPSGITDYGVVAKPINTSYGFSHHDLSKPAFVL 289  
 OY 202 PIP--YLLAVGSGDSSVRL--YDR-----RMLGTRATGNVAGRGTGMVARF-I 245  
 DB 290 PFPKEVAVAVRCSPLYRLRPIYNAEKNPPLISLPYMTIYAVATKNVFFYDTQOPVFAT 349  
 OY 246 PSHLNKSCRVTSLCYSEDQELIVSYSDYIYL--FDPKD--DTARELKT-----SAE 296  
 DB 350 VSNIHYS--RLNDLAWSSDGVTLIVSSTGYCLIFFEPELGDCEDEMETVLSVYLKSS 407  
 OY 297 ERREELRQPPVRLRLKGMSTGPRAPREBERDEQSPVSLMQRMSDMLSRMFEA 356  
 DB 408 ENATYVLRK--KROKRLKYSLDE-PR-KPEQESK-----PNT-----IRRA 444  
 OY 357 SEVAOSNRGRSRPRGSGSDISFLPYPSSEPLEVESETAMENVDPAGQIFQPSSTT 416  
 DB 445 SEAGITEVEPELDAENDSSTHSVSNKT--NSPKTKASEE--EKDTPAIRSRPKNST 500  
 OY 417 MSAQAHSTSPTESPASTLISLSPDEOROSVASGHTHTHOSDNNEKLSPKGTGPEV 476  
 DB 501 -----APMPAIRRAPRKP-----EDKXGSESKPKPD-----527  
 OY 477 LSLHSTEGTTSTIKLNTDEMSTLASSRGISGCKSEGESEFVPOSSVOP-----531  
 DB 528 -----EMEYTHRTQVOYKT-----VASPVKRKVSLEAVPAETSQPALAVIP 570  
 OY 532 -----EDSETKAPESSESDVTKYQEGVSAENPVENININITQSKFTAKPLDSSNGER 584

DB 571 VFEKEITSSDDKEESPCKSRPATPIYVRQOPRT-----GSSQFNTLPKSQPAKQA 623  
 OY 585 NDNLDRSCGV---PREES--ASSEKAREPSTDQSTSE--SATNENNTNPEQOFQTEATG 637  
 DB 624 TPVAVRRTPVLLIEMPVNPVPEAMAMPIDESTITPLPATVKKDSRKLPLPVTAKETK 683  
 OY 638 PSAHEFS-----TRSDALQDITDSDDDP 661  
 DB 684 PAVTEASEATCERTEDIRLYVEDYQEEETP 712

## RESULT 21

ID 095TD6 PRELIMINARY; PRT; 298 AA.  
 AC 095TD6;  
 DT 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE SD07783p.  
 GN CG5124.  
 OS *Drosophila melanogaster* (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Phryganea; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxId=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
 RA Nuno J., Pacled J., Paragas V., Park S., Phouanavong S., Wan K.,  
 RA Yu C., Lewis S.E., Rubin G.M., Celinker S.;  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: CONTAINS 2 WD REPEATS (TRP-ASP DOMAINS).  
 DR EMBL: AY059458; AAL13364.1; -  
 DR Flybase: FBgn0034305; CG5124.  
 DR InterPro: IPR000923; BlueCu\_1.  
 DR InterPro: IPR001440; TPR.  
 DR InterPro: IPR001680; WD40.  
 DR Pfam: PF00515; TPR; 3.  
 DR Pfam: PF00400; WD40; 2.  
 DR PROSITE: PS00196; COPPER\_BLUE; UNKNOWN\_1.  
 DR PROSITE: PS00678; WD\_REPEATS\_1; UNKNOWN\_1.  
 DR PROSITE: PS50082; WD\_REPEATS\_2; 1.  
 DR PROSITE: PS50294; WD\_REPEATS\_REGION; 1.  
 KW Repeat; WD repeat.  
 SQ SEQUENCE 298 AA; 33989 MW; C5A54507FF69A9A43 CRC64;

## Query Match

Best Local Similarity 4.7%; Score 212.5; DB 5; Length 298;  
 Matches 51; Conservative 24; Mismatches 39; Indels 13; Gaps 4;

OY 693 RKMEELDTNIRPLVKKYKGRNSRMVIXKANFNGA--NRYMGSGDCGHIFINDRHT 750  
 DB 157 RLKNAEYDLRSTARDYQR--YVGHCVTTDIKAKANYIGSGEETIAGSDGNNYIMWGDY 215  
 OY 751 AEHLMLLEADNHNVCLOPAPFPDPIILASGIDYDIKILSPLESRJPNRKLAEVITRNE 810  
 DB 216 GKIRAYRADSAIVNCVQPHPSICMLATSGIDINIKIMSCAAS-----AEE---RPN 265  
 OY 811 LMLEETR 817  
 DB 266 LVADVTR 272  
 RESULT 22  
 ID 09BIT5 PRELIMINARY; PRT; 747 AA.  
 AC 09BIT5;  
 DT 01-JUN-2001 (Tremblrel. 17, Created)  
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)  
 DE Chromatin assembly factor-1 p105 subunit.

CAFL-105 OR CG12892.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NCBI\_TaxID=7227;  
 RN  
 RP  
 SEQUENCE FROM N.A.  
 RA MEDLINE=21424640; PubMed=11533245;  
 RA Tyler J.K., Collins K.A., Prasad-Sinha J., Amiot E., Bulger M.,  
 RA Harte P.J., Kobayashi R., Kadonaga J.T.;  
 RT Interaction between the Drosophila Caf-1 and ASF1 Chromatin Assembly  
 RT Factors";  
 RL Mol. Cell. Biol. 21:6574-6584(2001).  
 CC -1 SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).  
 DR EMBL: AF367178; AAK31264.1; -;  
 DR FLYBASE: FBgn0033526; Caf1-105.  
 DR InterPro: IPR001632; GpoteIn\_B.  
 DR InterPro: IPR001680; WD40.  
 DR Pfam: PF00400; WD40; 5.  
 DR PRINTS: PR00319; GPOTEINB.  
 DR ProDom: PD000018; WD40; 1.  
 DR SMART: SM00320; WD40; 5.  
 DR PROSITE: PS00678; WD\_REPEATS\_1; UNKNOWN\_1.  
 DR PROSITE: PS50082; WD\_REPEATS\_2; 3.  
 DR PROSITE: PS50294; WD\_REPEATS\_REGION; 1.  
 KW Repeat; WD repeat.  
 SQ SEQUENCE 747 AA; 83406 MW; C7EBE839243B8767 CRC64;  
 Query Match 4.7%; Score 211.5; DB 5; Length 747;  
 Best Local Similarity 20.6%; Pred. No. 1.6e-05;  
 Matches 167; Conservative 103; Mismatches 282; Indels 257; Gaps 40;  
 QY 8 PHLLW-----DYKRSLSGLEPS--RLNSR-----YLGRREFIQRLKE--A 45  
 DB 6 PEISWMNRDPVLSVDIQNGIGLRSPTRICRIASGSDAHVLIWYVRSDAEVDVLEIA 65  
 QY 46 TLNVHDCVNTICMNDTGERIISGSDTKLVY-----SNPYSRVLT 88  
 DB 66 DLSRHRANAVNAWSPNGELIASGDSEVYFIMKORADHEVNIYDADGSEQDEKVMIL 125  
 QY 89 IR--SGHRANISAKPLPCTNPKOIVSGSDGVIFVTNVEQ----- 127  
 DB 126 LKYLKSHREDIYDLWAP--NSQFLVSGSVNDNTAMLVHSGSLAIIIDHKCYGVGVAM 183  
 QY 128 -----DAETNR-----OCQFTCHVGYEIMTVNDPYTF--LS 159  
 DB 184 DPCNOYIATMSTDRMRIFDANKRVLLHRSKC-----VLPVKEDHMHGSKMR 232  
 QY 160 CGEDGVKRFEDRIKTSCKRECK-----DDIL--IN-----CRAATVAICP 201  
 DB 233 LYQDGLQFFRRRL--CETPDGKLLLPSTGYDGVVKKPINTSYGFSKYDLKSPAVL 289  
 QY 202 PIP--YLLAVGSSSVSR--YDR-----BMLGTRANGYAGKGTGMVARF-I 245  
 DB 290 PEPNEAVAVKRCSPVYLRIPYNAEKNPILIPYMIYAVATKNAVFFPDIDQPPPAI 349  
 QY 246 PSHLNKSCRVTSLSYSEDGQELIVSYSDYTL--FDPRD--DTARELKT-----SAE 296  
 DB 350 VSNIHYS--RLNDLAWMSDGTVLIVSSTGYCSLITFEPELIDCYEDMETVLSVLKSS 407  
 QY 297 ERREELROPVVKRLKRCMSDTPPARPESEREDEGSPNVSLMQRMSDMLSRFEER 356  
 DB 408 ENATYVKK--KROKLKRVSLDE-PR-KPLQEKSK-----PNT-----IRRA 444  
 QY 357 SEVAOSNRGRGSRPRGSGQSDISLTPVPSPDLEVESETAMEVDTPAEOFIQSTST 416  
 DB 445 SEAGTIVEPELDAENDSTSHSVSNKT--NSPKTKASEE--EKPTPAIRSRPKKNST 500  
 QY 417 MSAQAHSTSPSTPSTLSSPDSEORQSVASGHTTHOSDNNEKLSPKGTGEPEY 476  
 DB 501 -----APWPIAIRAPRKP-----EDKNGSESSKRPD----- 527

QY 477 LSLHSTEGTTSTIKLNFTEWSSIASSRGICSGCKEGBQESRPVQSSVQP----- 531  
 DB 528 -----EMEYVHTKTTOVOKT-----VASVKKRVSTEAAPPAETSQPALAVIP 570  
 QY 532 -----EGDSEIKAPEESEDVTKQEGSVANPENHNINIQSPKTFAPKLDNSGER 584  
 DB 571 VEKEKLISSDKFESPEKSKRPATPIQVRQPTP-----GSSQFNLTPKSPAKQA 633  
 QY 585 NDNLNDRSGV-----PEES--ASSEKAKEPESDQTSYE-SATNENNTNEPOFQTEATG 637  
 DB 624 TPAVARTTRVLEEMVNSVPVMEAMAMPDEISITPLPATKKSRRKPLPTVTEFK 683  
 QY 638 PSAHEETS-----TRSDALQDTDDSDDDP 661  
 DB 684 PAVIEASEATCERTEDIRLYVEDTQETP 712  
 RESULT 23  
 ID Q9FWE0 PRELIMINARY; PRT; 131 AA.  
 AC Q9FWE0;  
 DT 01-MAR-2001 (TREMBLrel, 16, Created)  
 DT 01-MAR-2001 (TREMBLrel, 16, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel, 20, Last annotation update)  
 DE Putative WD-repeat containing protein.  
 GN OSJNB001511.22.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 NCBI\_TaxID=4530;  
 RN  
 RP  
 SEQUENCE FROM N.A.  
 RC STRAIN=CV. NIPPONBARE;  
 RA Buel C.R., Yuan Q., Moffat K.S., Hill J.N., Jenkins C.N., Burr P.C.,  
 RA Hsu J., Zismann V., Pal G., Bowman C.L., Fujii C.Y., Vanaken S.E.,  
 RA Bowman C.L., Craven B., Utterback T.R., Khalak H., Feldblyum T.V.,  
 RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.;  
 RT "Oryza sativa chromosome 10 BAC OSJNB001511 genomic sequence."  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1 SIMILARITY: CONTAINS 2 WD REPEATS (TRP-ASP DOMAINS).  
 DR EMBL: AC051633; AAG13587.1; -;  
 DR InterPro: IPR001680; WD40.  
 DR Pfam: PF00400; WD40; 2.  
 DR ProDom: PD000018; WD40; 1.  
 DR SMART: SM00320; WD40; 2.  
 DR PROSITE: PS50082; WD\_REPEATS\_2; 1.  
 DR PROSITE: PS50294; WD\_REPEATS\_REGION; 1.  
 KW Repeat; WD repeat.  
 SQ SEQUENCE 131 AA; 14750 MW; EA37C0E1C19A32F9 CRC64;  
 Query Match 4.6%; Score 207; DB 10; Length 131;  
 Best Local Similarity 35.5%; Pred. No. 3e-06;  
 Matches 44; Conservative 25; Mismatches 51; Indels 4; Gaps 1;  
 QY 1 MSKGSYF-----HLMDVKKRSISGLEDPKRLKRYIGRRFIORLKLEATLVNHDGCVNT 56  
 DB 1 MRPPWKPYPKARAGADLCIREVGLLPFRFARRAASEDLVNRQLVHRLNHTGCVNT 60  
 QY 57 ICNNDGEYILSGSDTKLVISNPKRYVLTTRSGHRANIFSAKPLPCTNDRQIVSCSG 116  
 DB 61 VGFNADGDTLISGSDQWALMDWDGAIKLPQPSGSHSNVFPARPPYNTDQITVTCAL 120  
 QY 117 DGYI 120  
 DB 121 DGEV 124  
 RESULT 24  
 ID O60336 PRELIMINARY; PRT; 1217 AA.  
 AC O60336;  
 DT 01-AUG-1998 (TREMBLrel, 07, Created)

DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE KIA00596 protein (Fragment).  
 GN KIA00596.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN;  
 RX MEDLINE=98290545; PubMed=9628581;  
 RA Nagase T., Ishikawa K., Miyajima N., Tanaka A., Kotani H., Nomura N.,  
 RA Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. IX.  
 RT The complete sequences of 100 new cDNA clones from brain which can  
 RT code for large proteins in vitro.";  
 RL DNA Res. 5:31-39(1998).  
 CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).  
 DR EMBL: AB011168; BAA25522.1; -;  
 DR Interpro: IPR002114; HPR\_Serp\_site.  
 DR Interpro: IPR001680; WD40.  
 DR Pfam: PF00400; WD40; 8.  
 DR PRINTS: PR00320; GPROTEINBPT.  
 DR SMART: SM00320; WD40; 7.  
 DR PROSITE: PS00589; PTS\_HPR\_SER; UNKNOWN\_1.  
 DR PROSITE: PS50082; WD\_REPEATS\_2; 1.  
 DR PROSITE: PS50294; WD\_REPEATS\_REGION; 1.  
 KW Repeat; WD repeat.  
 FT NON\_TER  
 SQ SEQUENCE 1217 AA; 131098 MW; 0BC4EAC6722BEF5 CRC64;  
 Query Match 4.5%; Score 204.5; DB 4; Length 1217;  
 Best Local Similarity 20.6%; Pred. No. 8,76-05;  
 Matches 187; Conservative 121; Mismatches 351; Indels 247; Gaps 43;  
 QY 39 ORLLEATLVNHDGCVNTICWN--DTGEYLISGSDPTKLV-----ISNPYRKVLTIRSG 92  
 DB 218 EMLKVEA-----HSEILCLEYSKPDITGLKLASNRDLHVLADAGRYTS---LQOTIDE 270  
 QY 93 HRANIFAKELPCTNDKOIYVSCSDGVIFETVNBQDAETNRCQFT-CHY---GTYEI 147  
 DB 271 HSSITAVKFAASDGOVMTISCGADKSIYFETAKSGD--GVQFTTHHVARKTTLAYDM 327  
 QY 148 MYVNDYFELSGEDGTVMFDRITKTSCKEDCK-----DILLINCRRAATVAILCP 202  
 DB 328 DVERSMKTYIAGC-QDNINILFN--ISSGKQKLFKSGSGEDGLIVQ-----TDP 376  
 QY 203 IPYLVAGCSDSSVRIYDRMLGTRATGNTAGRGTMVAFIPSHLNKSCRYTSLCYS 262  
 DB 377 SGYIATSCDKNLSTID-----FSSGECVATMTGH---SEIVTGKMF 417  
 QY 263 EDGQELIVATSSD-YIYLFPPKDDTAAELTPSAEERREE--LRQPPVKRLRG--- 314  
 DB 418 NDKKH-LIYSVGSCLFVWRLISEMTISMRORLAELRQROGKQGPSPQASGPNRH 476  
 QY 315 ---DMSDTPGRAPRESEREDG---EOSPVSIMQR-----MSDLSKWE 354  
 DB 477 QASMSLSPGALSDDSKEDDECTEDELPAFLAKSTKALASVPSPALPSLSHW-- 534  
 QY 355 EASEVAOS-----NKGGRSRPRG-----GTQSD 379  
 DB 535 EMERAEESVGLDPAPAPANGPRRRGRWVQGYELSVRSMILQRLTAPSLQDPQDS 594  
 QY 380 ISLTPVPSPDLEVTAEVTPAQFLQPTSSSTMSQAHSSTSPSPSPSTPLSS 439  
 DB 595 LALIPSGPRKHGQEALETSL-----TSQNEKPPRQASQPCSYPHIRLRL 641  
 QY 440 PDEEOROSVNAS-----GHHTHOSDN-----NNEKLSPK 469  
 DB 642 EEEGVFQDLLEPAILEDGIYVPEPSDNPMTSEFOVQAPARKGLGRYVPSRSSEKSP- 700  
 QY 470 PGTGEVPLSLHYSTEGTSTIKLNFTEWMSIASSSRIGSHCK--SEGQESFVQSS 527

DB 701 ----DSACSVYSS--SCLSSEPHTEDESESTPLSYVGISSDLPEAPAGDEEEEEE 754  
 QY 528 VQP---PEGDSETKAPEE--SESDVTKYQGVGAENPV---ENHINITOSDKF-----TKP 576  
 DB 755 MCPYLOQESPPOTPPQEOPLKHFETLASGAPAPGVQVPESESRSSISRLLOYQTP 814  
 QY 577 LDSNGERNLNL-DRSCGVPEESASSEK--AKPEPSTDTSTESANTENNTPQ-- 630  
 DB 815 LNEPSSSSILALMSRPQVQASGEQPRGNCANPPGAPPEVPESSG-----NSPQA 868  
 QY 631 -----FQETATGPASHEETSTRDSALQTDSDDDPVLIPARRA-----GPDR 676  
 DB 869 ASVLLPRLRLNDSSWAPRRVATASPFGLQKQSVHS--LVQERHESLQAPSPGAL 925  
 QY 677 RSAVARIOEFFRRKREKREMEELDTINIRPLVKWYKGH-----RNSRTWKANFWG 730  
 DB 926 LS-----KEIEAQDGLSIPPADGPPSRHSIQNTTSMATISSISYV 970  
 QY 731 ANFWSSGDCGHFTWDRHTAEHMLLEADNH--VNCLOPHFPFDILASSGIDYDIKIMS 789  
 DB 971 ENLGLVAPQAHAPI--RVSPLSKLALPSRAHLVLDIPKPLDREPLTAA-----FS 1019  
 QY 790 PLEESR 795  
 DB 1020 PVTGR 1025  
 RESULT 25  
 ID Q17490 PRELIMINARY; PRT; 5170 AA.  
 AC Q17490;  
 DT 01-NOV-1996 (TReMBLrel. 01, Created)  
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
 DE B0350.1 protein.  
 GN B0350.1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditioldea;  
 OC Rhabditiida; Pelodierinae; Caenorhabditis.  
 OX NCBI\_Taxid=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,  
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,  
 RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,  
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Woldman P.,  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans.";  
 RL Nature 368:32-38(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Gattung S.;  
 RT "The sequence of C. elegans cosmid B0350.";  
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Waterston R.;  
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U50071; AAA93447.1; -;  
 SQ SEQUENCE 5170 AA; 575848 MW; 27C2DBFAAE03A7B CRC64;  
 Query Match 4.5%; Score 202; DB 5; Length 5170;  
 Best Local Similarity 20.7%; Pred. No. 0.00089;  
 Matches 125; Conservative 84; Mismatches 248; Indels 146; Gaps 24;





DB 558 GGGST-----TSSDPVEASGNGN-----SSTOSSSTTTTSSDEQOTS-SSDPV 606  
 QY 593 CGVPEASSEKAKEPETSQDOTSATSATN-----ENNTNPEPQTEATGSAHEETSTRD 648  
 DB 607 SEVAGSSSSIGDGNSSGSSSTTTTSSDGGOSTSSDP--VEASGNGNSSTOS 664  
 QY 649 SALQDTDDSD-----DPLPGARYAPGDRRS 678  
 DB 665 SSSTTTTSSDEQOTSSDPV-VEVAGSSSSNGDGN 701

## RESULT 27

ID 096006 PRELIMINARY; PRT; 1183 AA.

AC 096006; 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE LD47622P.  
 GN CG7337.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NC NCBL\_taxid=7227;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=BERKELEY;  
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
 RA Champagne M., Chavez C., Dorsett V., Farfan D., Frise B., George R.,  
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
 RA Nuncio J., Pacle J., Paragosa V., Park S., Phouanavong S., Wan K.,  
 RA Yu C., Lewis S.E., Rubin G.M., Celnik S.,  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: CONTAINS 8 WD REPEATS (TRP-ASP DOMAINS).  
 DR EMBL: AY052025; AK93449.1; -  
 DR Flybase: FBgn0031374; CG7337.  
 DR Interpro: IPR001680; WD40.  
 DR Pfam: PF00400; WD40; 7.  
 DR PROSITE: PS50082; WD\_REPEATS\_2; 2.  
 DR PROSITE: PS50294; WD\_REPEATS\_REGION; 1.  
 KW Repeat; WD repeat.  
 SQ SEQUENCE 1183 AA; 128147 MW; 43B043CE7A870BCD CRC64;

Query Match 4.4%; Score 200; DB 5; Length 1183;

Best Local Similarity 20.3%; Pred. No. 0.00016;

Matches 153; Conservative 107; Mismatches 333; Indels 162; Gaps 28;

QY 41 LKLEATLNVHDCVNTICNDTGEY-----LSGSDDTKLVISNPYSRKVLYTIRSGHRA 95  
 DB 176 LRLTTTEAHES--EVLCEIYSNEKIERKLLASASRDRLHFDVADQNYLLQLTLDHSS 233  
 QY 96 NISAKFLPCTNDKOIVSCSGD-----GVIEFYNVBQDAETNQCQFTCHGYTYE 146  
 DB 234 SITSIKFGVAGLNFQMIISCADSIKFRSGNITFMNGTTSCT-----TLVD 282  
 QY 147 IMTPVNDPYTFLSCGEGDGVYRWEDTR-IKTSCTEKEDCKDILINCRRAATVAICPPIPY 205  
 DB 283 MEVDSNAKHLLTAC-QDRNRYVYGTQNAKQTKFKGSHSD-----EGSLIKSLDPSGI 335  
 QY 206 YLAVGSDSSVRYD-----RRLMGRATGNYAGGTTGWARFIPSHLNKSCVYTS 258  
 DB 336 YVATSTCDKLAVDYYSNECMARMG-----HSELVTG 369  
 QY 259 LCXSEGGQELIVSYSDYILFDPKDDTARELKTPSAEEERELR-QPPVKRLRLGDM 316  
 DB 370 LKFTNCRHLISAGGCGCIFIMQVPHDM---ITTMQARMGQRLRSRHAUPLPLADISP 426  
 QY 317 SDTGPRAPRSEERDEGSPNYSLMOQMSDM--LSRMFEASEVAQSNKGRGRSRPBG 374  
 DB 427 PDGIVLESPTSIEOP-QLOPKFGVAERFSDVQQLPQWARKKAADSDGSLSTPTPSGC 485

QY 375 TS-----OSDISTLPVPSSPDLEVSETPAMEVDTPEAQFLQPTSTSTMSAQHSTSSPTES 430  
 DB 486 SATVPMGMAHSAASMGNLSSPSQOMTGLAPRARGRMAQ-----RSTQLETRADDLRNSSES 539  
 QY 431 P-----HSTLLSSPDEQROSVASHHHHHOSDNN-NKLSKPKPTGEPVLSH 480  
 DB 540 PLGTSSVSGHSGVNVOTSDNSASSKDTITNYLLESDSSIDSGMTRRELAFIGSSN 599  
 QY 481 YSTEGTTTSTIKLNFTEWSSIA-SSSRGIGSHCKSEGESEFVQSSVOP-----530  
 DB 600 NGTVVTVSSV-----SSIAVSASNGAMSTSGGAQOQLPDKRKGLKGLFDFHTH 650  
 QY 531 -PEGDSETKAP-EESSEDVTKYQEGVSAENPEVNHINTOSDKFTAPPLDSNGERDLN 588  
 DB 651 DHDGDVEDISDGETRSSDHGMFNPLAPSPPTDKVTAMNDEL-RKSVRQKFEKSLQ 709  
 QY 589 LDRSC--GVPEASSEKAKEPETSQDOTSATSATNENTNPEPQTEATGSAHEETSTRD 646  
 DB 710 LTPSALSGNGSSHTASTGTSTDEDEGSTPSA-----ENAKSLASTLGSSSNLPQSS 764  
 QY 647 RDSALQDTDDSDDPVILPGARYAPG-----DRAS-----AVARIQE 685  
 DB 765 TNSFLH-----AALPEGGLTTPMERGSSSRKSIKHTENGKSYAAPPT 810  
 QY 686 FRRRRKRRKMEELDTNIRPPLVKKYKGRNSR 720  
 DB 811 ITRKYSTKREELQVINKYQOLENV--GHRPLR 843

## RESULT 28

ID 09V054 PRELIMINARY; PRT; 2351 AA.

AC 09V054; 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE CG7337 protein.  
 GN CG7337.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NC NCBL\_taxid=7227;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=BERKELEY;  
 RA MEDLINE: 20196006; PubMed-10731132;

RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Y., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Bailew R.M., Basu A.V., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,  
 RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garcia N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris J.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegian C.,  
 RA Jialili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mishina N.V., Moberly C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spelling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svitkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000)  
 CC -1- SIMILARITY: CONTAINS 11 WD REPEATS (TRP-ASP DOMAINS).  
 DR EMBL: AE003584; AAF51327.1; -.  
 DR Flybase: FBgn0031374; CG3337.  
 DR InterPro: IPR001680; WD40.  
 DR Pfam: PF00400; WD40; 10.  
 DR SMART: SM00320; WD40; 9.  
 DR PROSITE: PS00082; WD\_REPEATS\_2; 2.  
 DR PROSITE: PS50294; WD\_REPEATS\_REGION; 1.  
 KW Repeat.  
 SQ SEQUENCE 2351 AA; 258254 MW; AE1402C714709AA8 CRC64;

Query Match 4.4%; Score 200; DB 5; Length 2351;  
 Best Local Similarity 20.3%; Pred. No. 0.00042;  
 Matches 153; Conservative 107; Mismatches 333; Indels 162; Gaps 28;

OY 41 LKLEATLVNHDGCVNTICWMDTGEYI-----LSGSDTKLVISNPYSRKVLTTIRSGHRA 95  
 DB 501 LRLTTIEAHES--EVLCELEYSEKIERKILASARDRLIHVDVANOYLLQTLDDHSS 558  
 OY 96 NIFSAFLPCTNDKQIVSCGD-----GVITYTNEQDAETNRCCOFTCHYTYE 146  
 DB 559 STISIFVAGINLFQMIISGADKSIMFRSFGNIFMKGTSGKT-----TLXD 607  
 OY 147 IMVVPDPTFLSGEDGYRWFDR-IKTSCKECKDKDILLNCRRAATVAICPPIPY 205  
 DB 608 MEYDSNAKILTRC-ODRNRVRYGTQNAQTKFKGSHSD-----EGSLIKSLDPSGI 660  
 OY 206 YLAVGCSDSVRIYD-----RRMLGTRATGNYAGRTTGVARFPLSHLNKSGVTS 258  
 DB 661 YVATSTGDKTALAYDYYSNCRMARWG-----HSELVYG 694  
 OY 259 LCYSEGGELIVSYSSDIYILFPKDDTARELTPSAERRELR--QPPVKRLRLRGDW 316  
 DB 695 LKFTNCRRLISAGDGCIFIQVPHDM--IYMQARMSQQLRSGHAPLPPLAPISP 751  
 OY 317 SDTPRARESERERDEGSPNYSLMQRMSD--LSRMPEAEVAQSNRGRSRPRG 374  
 DB 752 PDGIYESTPTEIOP-QIQPKFVAERISDVGQLPQMARAKAADSGLSIPPSGG 810  
 OY 375 TS-----QSDISTLPVSSPDLEVSSETAMEVDTPAQFLOPSTSTMSAQHSTSPTES 430  
 DB 811 SATVPMHAASSKGNLSSSPSQOMTGLAPRARGMAQ-----RSQULEADADLRNSSES 864  
 OY 431 P-----HSTPLISPDSEOROSEASGHTTHQSDNN-NEKLSPKFGTGEPLVSLH 480  
 DB 865 PLGTAVSSVGHSGVNYQTSDYNSASKDITYNQTYLESSESDISGEMTRGELKFGSSN 924  
 OY 481 YSTEGTTFSTIKLNTDEMSTIA-SSSRGIGSHCKSEGOESPVPSSVOP-----530  
 DB 925 NGTVAVVSSV-----SSTAVASNGAMSTGSGAAQORQLPKPKRLPGLRPDTHTH 975  
 OY 531 -PEGDEETAP--EESSEDYTKYQEGVSAENPVENHINITQSDKFTAKPLDSNGERDNLN 588  
 DB 976 DHDGDVEDISDGRISSDGMFYNNLAPSTPDPFKYLANEDL-RKSVRQKFEKSGGLQ 1034  
 OY 589 LDRSC--GYPEASASEKAKPEPTSDQSTESATNNNTNPEPQFTEATGPAHEETST 646  
 DB 1035 LPPSALSGMSSHTASTGTSTDEGSTPSA-----ENAEKSLASTLGGSSENLPSS 1089  
 OY 647 ROSALQDTDDSDDDPVLIGARHAPG-----DRRS-----AVARIQE 685

DB 1090 TNSFLH-----ALPEPGGLTTPMERGSSRRSISAKHNTENGSAVAAPT 1135  
 OY 686 FFRRRKREKMEELDTLIRRPVVKWYKGRNSR 720  
 DB 1136 ITKSTSTKKEELLQYINKKQGLENV--GHRRLR 1168

RESULT 29  
 Q17343  
 ID 017343 PRELIMINARY; PRT; 6994 AA.  
 AC 017343;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-JUN-1998 (TrEMBLrel. 05, last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, last annotation update)  
 DE UNC-44 ankyrins.  
 GN UNC-44.  
 OS *Caenorhabditis elegans*.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Peloderae; Caenorhabditis.  
 OX NCBI\_Taxid=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=N2;  
 RX MEDLINE=95263663; PubMed-7744957;  
 RA Otsuka A.J., Franco R., Yang B., Shim K., Tang L.Z., Zhang Y.,  
 RA Boontrakulpoontawe P., Jeyaprakash A., Hedgecock E., Wheaton V.I.,  
 RA Sobery A.;  
 RT "An ankyrin-related gene (unc-44) is necessary for proper axonal  
 RT guidance in *Caenorhabditis elegans*.";  
 RL J. Cell Biol. 129:1081-1092(1995).  
 RN [2]  
 RP REVISIONS, AND SEQUENCE OF 6126-6994 FROM N.A.  
 RC STRAIN=N2;  
 RA Otsuka A.J.;  
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U39847; AAB41827.1; -.  
 DR EMBL: U21733; AAB384.1; -.  
 DR HSSP: P42773; 11HB.  
 DR InterPro: IPR002110; ANK.  
 DR InterPro: IPR000488; Death.  
 DR InterPro: IPR001360; GH.1.  
 DR InterPro: IPR002383; GLA\_blood.  
 DR InterPro: IPR000906; ZUS.  
 DR Pfam: PF00023; ank; 24.  
 DR Pfam: PF00531; death; 1.  
 DR Pfam: PF00791; ZUS; 1.  
 DR PRINTS: PRO1415; ANKYRIN.  
 DR PRINTS: PRO0001; GLABLOOD.  
 DR SMART: SM00248; ANK; 21.  
 DR SMART: SM00005; DEATH; 1.  
 DR SMART: SM00218; ZUS; 1.  
 DR PROSITE: PS50088; ANK\_REPEAT; 22.  
 DR PROSITE: PS50297; ANK\_REPEAT\_REGION; 1.  
 DR PROSITE: PS500572; DEATH\_DOMAIN; 1.  
 DR PROSITE: PS005172; GLYCOSTYL\_HYDROL\_FL\_1; UNKNOWN\_2.  
 KW ANK repeat; Repeat.  
 SQ SEQUENCE 6994 AA; 775364 MW; 90CB449925D9923D CRC64;

Query Match 4.4%; Score 200; DB 5; Length 6994;  
 Best Local Similarity 20.6%; Pred. No. 0.0018;  
 Matches 124; Conservative 85; Mismatches 248; Indels 146; Gaps 24;

OY 191 RRAATVAICPPPIYLAAGCSDSVRIYDRMLGTRATGNYAGRTGVARFIPSHLN 250  
 DB 2344 RPISESPVASERPSHIAITTTTTRVYQDEKADQSSARES-----VRLVSEHE 2396  
 OY 251 NKSCHVTSICYSBEGQELIVSYSDIYILFPKDDTA--RELKTPSAER-----REE 301  
 DB 2397 HPACDESRLESEPAQSEPEVTHAESQFSNLYETTTTAAVREFYDDEQOASRATKER 2456  
 OY 302 LRPPPKRLRLRGDWSDTGRAPR--ESERERDEGSPNYSLMQRMSDMLSRKFE--AS 357  
 DB 2457 IEOSPASER--SIVSTERQSPQESDSELEPTSEKVNHTVETTTTTRVYRECPEPIAS 2513

```

QY 358 EVAQNR--GKRSRPRGTSQSDISTLPVPSPP--DLEVSEIAMEYDTPAEOFIQDSTS 414
DB 2514 ELEHASREFEFGSGNDRRSVPYDPAQEOYDVPSPAPSSHAESOPVESPPIHIVETTT 2573
QY 415 STMSAASHSTSSP--TESPHSTPLISSPDSEORQ-----SVASGHHTHHOSNNN 463
DB 2574 TTVTRERODEEYRPPSP--AEIPIPPSSHQSEBPIVETTTTTVTRELDIDEPKGN 2631
QY 464 EKLSPRKGT-----GEPVLS--LHYSTGTTSTIKLFTDEWSSIASSSRGIGS 511
DB 2632 VTPSPAPSSHAESRQVPEPVSHQRYPHVETTTTTVTNTSNLYDDEDVNPSEDPATQ 2691
QY 512 HCSSEGESEFVPOS--STQPEGDS-----535
DB 2692 HFQ---QSETSVHRSHSDSVESDGEIGSKVIGFAKKAGVAGVAAAPVLAAGAKA 2748
QY 536 -----EKAPAESEEDVTAKYOE--GVSANVENHINTQSDKFTAK 575
DB 2749 AYAPAEKDEDEDETSHPESPVPYQSE---QYODDSAQSSHDFEHNM-----2794
QY 576 PLDSNGERNDLMDRSCGVPESSASEKAKEPE--TSDQSTESATNENNTNPE-----628
DB 2795 PESPHEKEETEEEDHS--HPESPVSEKERRDQVSETTTTTVTREYNDEPDEQEHQ 2852
QY 629 -----PQOTATGPAHETSTRDSALQDTSDDPVLIPGARYRAGGDRSAVAR 682
DB 2853 GPHSPAPSSHTAEHPHIVETTTTTVTTRFQEEPER-----LEKQEDNSKSPSSH 2905
QY 683 IOE 685
DB 2906 SQE 2908

RESULT 30
ID Q23330 PRELIMINARY; PRT; 493 AA.
AC Q23330;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE Hypothetical 53.6 kDa protein.
GN ZC449.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Latreille P., Gattung S.;
RT "The sequence of C. elegans cosmid ZC449.";
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Watson R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: U41510; AAK39385.1; -
KW Hypothetical protein.
SQ SEQUENCE 493 AA; 53643 MW; 6EDC2FEA3CCCA390 CRC64;
Query Match 4.4%; Score 199; DB 5; Length 493;
Best Local Similarity 19.3%; Pred. No. 5.9e-05;
Matches 102; Conservative 80; Mismatches 219; Indels 128; Gaps 16;

```

```

QY 249 LNKSCAVTSLCYSEDOELIVSSDYITLFDPKDPTAREK--TPSARERRELKQPP 306
DB 59 LKNNKCLLTR-----ARRHLSTPATADDSEETTT 91
QY 307 VKRLRLGDMSDTCGRAPRAPERERDEQSPNVSLQMSDMLSRWEEA--SEVAQNRG 365
DB 92 VKSHKSKSTKTPADPKTEEPATSKKSEKKTEKKAVKTTTAAPEASAEETEPSEEN 151
QY 366 RGRSRPRGTSQSDISTLPVPSPPDLEVSEIAMEYDTPAEOFIQDSTSMSAASHST 425
DB 152 KTEESGNVSAEHTTTTSTEVSEBEOYVKSTEASTEEK-----STEASSEATTT 205
QY 426 SPTESPHSTPLISSPDSEORQSVASGHTHHOQDNNEKLSPKPGTGEPLSLHYSTEG 485
DB 206 SETSE-----TTESESE-----EETTPA-----KTHSKKT 233
QY 486 TTTSTIKLFTDEWSSIASSSRGIGSHCKSEGESEFVPOSVPPEGSEETKAPESSE 545
DB 234 TTTT-----EASTKDKSK--SHKTKKEKVTTPPESTESSSPSTSEPPQSE 283
QY 546 DVTYQEGV--SAENPVENHINTQSDKFTAKPLDSNGERNDLMDRSCGVPESSASSE 603
DB 284 STPSEESTEQSTERP-----KKEDKKDKKKKKKKED-----SEDDDDK 327
QY 604 KAKREPISDOTSTESATNENNTNPEPQOTATGPAHETSTRDSALQDTSDDPVL 663
DB 328 KSKSSSSSDSTDEKSTESDSSSEASEKTEVEGK-----KTLFSGPDEDDDD--379
QY 664 IPGARVRAGPDRRSAVARIOEFFRRKERRK---EMELDTLNTIRPLVKWYKGRNSR 720
DB 380 -----EGAGADAFSEKQATAPALSTTTTMAVSKKKKSESGFNIP 425
QY 721 TMKEANFWGANFYWGSDCGHITWDRHIAHMLLEADNHVYNCLOP 769
DB 426 LVI-----AGVFGVGLVAGVYFNNKKERENI,STVDERKELTTIAP 467

RESULT 31
ID Q23587 PRELIMINARY; PRT; 3507 AA.
AC Q23587;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE ZK783.1.
GN ZK783.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kersey J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shawken R.,
RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sprat J., Woldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Favello A., Vaudin M.;
RT "The sequence of C. elegans cosmid ZK783.";

```

RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.  
 RN [3] SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Waterston R.;  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U13646; AAC24418.1; -.  
 DR HSPB; P00736; IAPQ.  
 DR Interpro: IPR000152; Asx\_hydroxyl.  
 DR Interpro: IPR000561; EGF-like.  
 DR Interpro: IPR001881; EGF Ca.  
 DR Interpro: IPR001507; Endoglin/CD105.  
 DR Interpro: IPR001254; Ser\_protease\_inh.  
 DR Pfam: PF00008; EGF\_15.  
 DR SMART: SM00179; EGF\_CA; 12.  
 DR SMART: SM00001; EGF\_16.  
 DR SMART: SM00241; ZP\_1.  
 DR PROSITE: PS00022; ASX\_HYDROXYL; 9.  
 DR PROSITE: PS00022; EGF\_1; UNKNOWN\_1.  
 DR PROSITE: PS01186; EGF\_2; 22.  
 DR PROSITE: PS01187; EGF\_CA; 13.  
 DR PROSITE: PS00135; TRYPsin\_SER; UNKNOWN\_1.  
 DR Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.  
 KW SEQUENCE 3507 AA; 365315 MW; 154F0B687874D9DF CRC64;  
 SQ

Query Match 4.4%; Score 198; DB 5; Length 3507;  
 Best Local Similarity 20.6%; Pred. No. 0.00096;  
 Matches 160; Conservative 112; Mismatches 282; Indels 222; Gaps 33;

QY 77 ISNPKRVLTTRSGHRANIFSAFLPTNDKQIVSGSGVGFYINVEDQAFNQCQ 136  
 DB 1994 ISSSTSTKMDTSSKSPENVMSSSPVSSSTASSETV-----SPTSESSSSSEAP 2049  
 QY 137 FTCHGYTITIMVYNDPYFLSCGEDYVRFEDTRIKTSCTKEDCDILINCRAATS 196  
 DB 2050 LTSSPATTEVIT-----ESSVSKTTPKESSSEITVLSKSPE 2089  
 QY 197 VAIICPPIYLAIVGSSSVRIYDRMLGTRANGVAGRGITGVAFIFSHLNKRCRY 256  
 DB 2090 V-----TESSK-----SSSTPTSSQSVTSTVPTSKSTVLSSEAPV 2128  
 QY 257 TSL-----CYSDQOEILIVSYSDYIYLPDKDTPARELTPSAEERRE----- 300  
 DB 2129 TSTSPTEVHTSSEFKPGLSASTGTGDTNSTPTSSLASVKSASAPGTASAVPVKLS 2188  
 QY 301 ---ELNQPPYKRL-----RLRGDMDTGPRAPESERE-----RDGEQSPNVSLMQ 343  
 DB 2189 LSPDVQSPSTKTPDASTSTVQASSTSVKSTSEPSHVTKLITSSNPSSSVPTS 2248  
 QY 344 RMSDMLSRWFEASVAVQSNRGRGRS-RPRGGSQ-----SDISTLPTVP 387  
 DB 2249 PKS---TPVPESTEQPTSTTPSGQSLTPMANSSEVLTSEPHVLSLSPDVQSSTTP 2305  
 QY 388 S-----SPDLFVETAMEVD-----TPAQPLQ-----STSTMSQAHS-- 423  
 DB 2306 NNLSESVYETPKTSSVLSINSEPTTEAPTLSPDILSTTNLNSQSTSTVEDRSEI 2365  
 QY 424 TSPSTSPHSTPL-----SSPD---SEQQSVASGHHTHQSDNNKEL---SP 468  
 DB 2366 SSNSKSTPAPLIVSYTHVAVSSPDVPTSESPDILGSSSTENPEASKQTLISSTP 2425  
 QY 469 KPQT---GEPVLSLHYSTEGTSTSTL-KLANTDEWS---TASSRGIAGCKSGRQ 519  
 DB 2426 TPDITASEPEKSTKSPDLSSTSNVLSSTTPPESSSKSPVSSSTEGISVYTSF---E 2482  
 QY 520 ESTVPOSSVQ-----PPGDSSTKAPESSESDVTK-----YQGVSAEN- 558  
 DB 2483 FSVVPESTTSSVLEDLTKTPSPILFETTTASETSEPLEDLVAVSRHIELTSSENV 2542  
 QY 559 PVENHNINIOSDKFTAKPLDLSNGERNDLNDKSGVPESSASSEKAKEE----- 609  
 DB 2543 PRESESTTSSS--SSKPSQEPAG-----ILSTVAVPTSSVSLITASELTATSNTPPK 2595

QY 610 -----TSDQSTESATNENN-TNEPOFOEATGPSA-----HEESTRDSAL--- 651  
 DB 2596 QGRFTTSPKSLVASTSPSVTSSEPSSESTKRTVSTVSTTPTTESTTSESLTLTA 2655  
 QY 652 -----QPTDSDDDPVLIPGARYRAGGRDRSAVARIOEFFRRRKRKEMELDT 701  
 DB 2656 APSKPTSTESSEAP-TPAKTSEKPSNVST-----SKRSTENVET 2698

## RESULT 32

ID Q9DGL1 PRELIMINARY; PRT; 791 AA.  
 AC Q9DGL1;  
 DT 01-MAR-2001 (Tremblrel, 16, Created)  
 DT 01-MAR-2001 (Tremblrel, 16, Last sequence update)  
 DT 01-JUN-2001 (Tremblrel, 17, Last annotation update)  
 OS Retinitis pigmentosa GTPase regulator-like protein (Fragment).  
 OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;  
 OC Tetraodontidae; Takifugu.  
 OX NCBI\_Taxid=31033;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Vervoort R., Lennon A., Bird A.C., Tulloch B., Axton R., Milano M.G.,  
 RA Meindl A., Meindler T., Ciccodicola A., Wright A.F.,  
 RT "Mutational hot spot within a new RGR exon in X-linked retinitis  
 RT pigmentosa".  
 RL Nat. Genet. 0:0-0(2000).  
 DR EMBL: AF286475; AAC00554.1; -.  
 FT NON\_TER 1  
 SQ SEQUENCE 791 AA; 89672 MW; DA40A0FA9E3E39 CRC64;

Query Match 4.3%; Score 195.5; DB 13; Length 791;  
 Best Local Similarity 19.9%; Pred. No. 0.00019;  
 Matches 122; Conservative 90; Mismatches 294; Indels 107; Gaps 17;

QY 163 DGIYRWEDTRIKTSCTEDCKDILIN-----CRAATSVAIICPPIYLAIVGC 211  
 DB 231 EGSGRLE---KSSSPEDSGSVLRAQSESEAHGAFSKNSANVINIOPASSSEEGS 286  
 QY 212 SDSVRIYDRMLGTRATGNAG-----KGTGVANFIPSHLNKSCRYTSLCYSDQ 266  
 DB 287 QDKSEY-----TANDEEKESIQGTEEEEEEEDQKEQKTSADVSEEE 336  
 QY 267 EILVYS-----SDIYLFDPKDPARELTPSAEEREELRQPPVYKRLRGD 315  
 DB 337 ENYQSSDKASDEKEDSDTLH---PEDSQSVDEDEBEGESKSEQEEEEEED 393  
 QY 316 WSDTPAPARESERERDGEQSPNVSLMQRMSDMLSRWFE---ASEVAQSNRGRSRPR 372  
 DB 394 RSTP---AESESEDEKKGRK-----ERDVEEDEGESEVEESEGSDHSER 437  
 QY 373 GGTQSQD-----ISLPIYVSSPDLEFVSTAEVDTPAQFLOPTSSSTMSQAHS 422  
 DB 438 KKAASEEEEBEEDADSLSEEBGDSKDAEESDAESETGEGAEEDQSTERKESNEE 497  
 QY 423 STSPSTSPHSTPLSSPDSEORQSVASGHHTHQSDNNKELSPKPGTGEPLSLHYS 482  
 DB 498 EQSTKEEENNEEQSTKEEESNEEQSS--TKKEESNEEDSS-----TEREES 547  
 QY 483 TEGTTTITLNTDQESSLAASSRGIAGCKSGRQESFVPOSSVOPREGDSTKAPAE 542  
 DB 548 NEEQSTKEEESNEEQSTKEEESNEEESNEEESNEEESNEEEDQDEEETGEE 607  
 QY 543 SSEDVTKYQGVSAENPVENHNINIOSDKFTAKPLDLSNGERNDLNDKSGVPESSASS 602  
 DB 608 DEEEERSEQONDEAEEDDEEDIVQEEEEEETKEKEDE-----EEETEK 657  
 QY 603 EKAKEPESDQSTESATNENNTPPEFOFTEATGPSAHESTRSDALQDSDDDVP 662  
 DB 658 EEDDEEETDRVVEEDDAEENNEEESNEEESNEEESNEEESNEEEDDEESDEDEEESSE 717



RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amandides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Mortan J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champs M., Pfeiffer B.D.,  
 RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Adair J.F., Agayani A., An H.-J., Andrews-Plamkoc H., Baldwin D.,  
 RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Botshakov S.,  
 RA Borova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
 RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew J., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegami C.,  
 RA Jalali M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasro P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Matzel B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merulov G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nisskern D.R., Pacle J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reibert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Sprelding A.C., Stapleton M., Strong R., Sun E.,  
 RA Switskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT \*The genome sequence of *Drosophila melanogaster*.  
 RL Science 287:2185-2195(2000).  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A  
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 DR EMBL: AE003765; AAF56794.2; -  
 DR EMBL: AE003765; AAF56795.2; -  
 DR HSSP: P12111; 2KNT.  
 DR Flybase: FBgn0003137; Ppn.  
 DR InterPro: IPR000361; EGF-like.  
 DR InterPro: IPR003598; Ig\_C2.  
 DR InterPro: IPR003006; Ig\_MHC.  
 DR InterPro: IPR002223; Kunitz\_BPTI.  
 DR InterPro: IPR000884; TSP1.  
 DR InterPro: IPR002221; WAP.  
 DR Pfam: PF00047; Ig\_3.  
 DR Pfam: PF00014; Kunitz\_BPTI; 12.  
 DR Pfam: PF00090; tsp\_1; 5.  
 DR Pfam: PF00095; WAP; 1.  
 DR PRINTS: PR00759; BASICTPASE.  
 DR ProDom: PD000222; Kunitz\_BPTI; 12.  
 DR SMART: SM00408; IgC2; 3.  
 DR SMART: SM00131; KU; 12.  
 DR SMART: SM00209; TSP1; 7.  
 DR PROSITE: PS00217; WAP; 1.  
 DR PROSITE: PS00280; BPTI\_KUNITZ\_1; 11.  
 DR PROSITE: PS0279; BPTI\_KUNITZ\_2; 12.  
 DR PROSITE: PS00022; EGF\_1; UNKNOWN\_1.  
 DR PROSITE: PS00092; TSP1; 3.  
 KW Alternative splicing: Immunoglobulin domain;  
 KW Serine protease inhibitor.  
 FT VARSPLIC 2803 2803 L -> SVSVV (IN SHORT ISOFORM).  
 FT VARSPLIC 2844 2854 ENFKTWDSGI -> VASPLHPNAV (IN SHORT  
 FT VARSPLIC 2855 3060 ISOFORM).  
 FT VARSPLIC 3060 3060 MISSING (IN SHORT ISOFORM).  
 SO SEQUENCE 3060 AA; 331579 MW; ACA31D3E558C7C0 CRC64;

Query Match 4.3%; Score 194.5; DB 5; Length 3060;  
 Best Local Similarity 20.4%; Prid. No. 0.0014;  
 Matches 96; Conservative 77; Mismatches 171; Indels 127; Gaps 18;  
 QY 259 LCYSEDOGLVYSSSYITLFPKDDTA---RELKPSAEERRELOPPVRLRLRGD 315  
 DB 562 ICEPFDKTV-----PPADDSSCKNKEFKPSODEGEERKVP-----GE 602  
 QY 316 WSDTGP-----RAPERER----- 329  
 DB 603 WF-TGPFGRKSKPGGGERVYLCLNGTKSVNCDKEKPELSEKNSACTEDELPL 661  
 QY 330 -----ERDEGSPNVSLWQRMDSMLRWFEASEVAQSNRGRSRPGTQSQDSTL 383  
 DB 662 TSTDKPDEDEDCDEDEGILLISGLSD--DEKSEYIDLEGAKTE----- 706  
 QY 384 PTYPSSEDLVETAMVDPAPQFLQPSSTMSNAHSTSPTEPHSTPLSPDSE 443  
 DB 707 -TPPEADLMQSDS---PPYDFE--ESTGTPEGSGYDSESTPDSGISTE--GSGDDE 757  
 QY 444 Q-----RQSYEASGHTHHOSDNNEKLSPPGGEVPLSLHYEGTTSTKLNF 495  
 DB 758 ETSEASTDLSSTDSSTDSSTDSSTDSSTDSSTDSSTDSSTDSSTDSSTDSST 812  
 QY 496 TDWSSIASSSGRGISGCKEGEGESFVQSSVQPPGDSERKAPESSEDVTKYQGV 555  
 DB 813 -ETGVSSTSTVSSSTESASSESTVSGASDSTGTSNADSTPESTASSTDDSTD 870  
 QY 556 AENPVENHITQSDKRTAPLDSNGERDLNDRCGVPEESASEKAPETSGT- 614  
 DB 871 SS---DNSSVSSSTESASSTSVSDSSDSDGSTD--GVSTTENSDDSTSDST 924  
 QY 615 ---STESATNNTNPEPOPATGPAHESTRD--SALDPTDSDDD 660  
 DB 925 SSDSTDS-TSDQTETPESTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 974  
 RESULT 35  
 ID 090A02 PRELIMINARY; PRT; 1680 AA.  
 AC 090A02;  
 DT 01-MAY-2000 (TREMREL. 13, Created)  
 DT 01-MAR-2002 (TREMREL. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMREL. 21, Last annotation update)  
 DE Hypothetical 184.6 kDa protein.  
 GN H24G06.1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Rhabditidae; Rhabditidae; Rhabditidae;  
 OC NCB1\_Taxid=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA None.  
 RT "Genome sequence of the nematode *C. elegans*: a platform for  
 RT investigating biology. The *C. elegans* sequencing Consortium."  
 RL Science 282:2012-2018(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Ryan E., Delehaunty A.;  
 RT "The sequence of *C. elegans* cosmid H24G06."  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Waterston R.;  
 RT "Direct Submission."  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF101309; AL32229.1; -  
 DR InterPro: IPR001680; WD40.  
 DR Pfam: PF00400; WD40; 10.

DR SMART; SM00320; WD40; 11.  
DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN\_1.  
DR PROSITE; PS00682; WD\_REPEATS\_2; 2.  
DR PROSITE; PS0294; WD\_REPEATS\_REGION; 2.  
KW Hypothetical protein.  
SQ SEQUENCE 1680 AA; 184618 MW; 85C6105233DE0FDA CRC64;

Query Match 4.38; Score 194; DB 5; Length 1680;  
Best Local Similarity 22.18; Pred. No. 0.00065;  
Matches 203; Conservative 119; Mismatches 402; Indels 194; Gaps 43;

QY 47 LNVHDCVNTICWNNICE-----YILSGSDPTKLVIN--PYSKR-VLTTIRSGHRA 95  
DB 512 INAHDDVYIDSDHASTSSHDPIPLASGGRDFYHARRRIRYSSQFHCAYVDHQS 571  
QY 96 NISFAELPCTNDKQI--VCSGSGVIFYTNEQDAETNRKOCFT---CHYGTIEIM 148  
DB 572 AIKSIKF--ASNNGQLHLTYAASDRSLIIV--KLSSFSDHCEFTVQIMSVASISIGDM 626  
QY 149 TVPNDPTFLSCGDEGIVRFDRIRKTSCKEDCKDILLNCRANISVAICPIPYLA 208  
DB 627 NFKHVDLFVYAGHDRLRQFDINGKTVREVKGTDDVAHSGK--ILKVAIDHSGSYAIS 684  
QY 209 VCGSDSVRIYDRMLCTRATGNVAGRTGMVARFIPSHLNKNSCHVTSICVSEGOEI 268  
DB 685 V-CSDKFTVYTDLR-----SGVCL---AVLCGGAATDPTSDDFKNV 724  
QY 269 LVSYSSDIYLFDP-PKDDIARELKTPSAER-REELRQPYKRLRLGDWS---DTGP 321  
DB 725 IVTTSNCSIFITWOLAKMLTERMI--SAQVRLMEVMTATPDSLIGSGSETISGDSNS 781  
QY 322 RARPEREREDQSPVAVSLMQRMSDMLSRWFEASVAOSNRGRGSRPRGG----- 374  
DB 782 FGRLPGLAPESGS--SASLYSDDDDS---TRFSSSVRSRKL-ILPGLGLIDSS 832  
QY 375 -TQSQDSTLPYVPSPDLE-----VSETAMEVDTPAEQFLQPSSTSSAQAHSSTSS 426  
DB 833 YAVGDSSEFPAVQSAVAERRTTNLFSDQYETDVSERQ--SDYVSRKKTILFDED 890  
QY 427 PTESPHSTPLISPDSEBORQSEASGHHTH---HOSDNNEKLSPKPGTG----- 473  
DB 891 QDSNLGSAQVLAAPINEDRRSASPLYPDQQLGQYOSKSMMLRVYTGCGVAVQTAKE 950  
QY 474 ---EPVSLHSTYGTGTTSTIKLNFDP--EWSSTIASSSRGIGSHCKS--EQQESFPVQS 526  
DB 951 LMSQSLASQNRNSQSGSSSHLSTNTSGRMWGMDPQOSNNNDWHPSTVDIHVSPVAT 1010  
QY 527 SVQPE-----GDSETKAPEESSED---VTKYQGVSAENVEVNI--NITQSDKPTA 574  
DB 1011 SMIPPOHOGYGRDHTSTIPMPDSTHPPLAPRTSRVLTASQALQOIQASSPFR 1070  
QY 575 KPLDSNGERNDL-----NIDRSCGVPEESASSEKAKEPETSDQSTESATN 621  
DB 1071 KSMORNLISRLFLNGGAQOPTVWSPALANOGAPRRSNSNLFAATNLEVPPTSTN--LS 1128  
QY 622 ENNTNPPQ---YQTEATGSAHEHSTRDSALOD--TDDSDDDVYLIP--GARRAGAGD 675  
DB 1129 RRHTDKOPKRLFTT-----TVQTRTSLDDEVESDDDAITSTFEPEKKNRIGY 1179  
QY 676 RRSVAARLQ---EPRRR--KERKMEELDTINIRPL-----VKMYKKGRNSR- 720  
DB 1180 RRSSTVYVDRADAVARRVAGTSPKSDRADLNPASSLSRSQSPKFLSOMINRD 1239  
QY 721 TMKEANFWGANFY-----MSGSDCGHIFIMDRHTAEHMLLEADNHVYVNCIQPHFD 773  
DB 1240 TPSSPSSSTGSIYGRNLQORRGSDVSVYASATRGDEMKSIDALNKL----- 1289  
QY 774 PILASSGIDVDIKIWSPLEBSRIRNKLIADLV-ITRNEMLMEERNTTIVASFTMLRLA 832  
DB 1290 -----MAVRSKIHQSEENIRKSTENLALIKN--LEEASNHSTPRA---RTSR 1331  
QY 833 SLNHIRADRLGDSRSGS 850  
| | | | |

DB 1332 SSNLRNADNLGNMESFCT 1349

RESULT 36

ID Q9P785 PRELIMINARY; PRT; 451 AA.

AC Q9P785;

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE Hypothetical serine-rich repeat protein.

GN SPBC1711.05.

OS Schizosaccharomyces pombe (fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC NCBI\_TaxID:4896;

RN 11

RP SEQUENCE FROM N.A.

RC STRAIN=972H-;

RA Aert R., Volckaert G., Wood V., Rajandream M.A., Barrell B.G.;

RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL353012; CAB88235.1;

SQ SEQUENCE 451 AA; 47427 MW; 36F254388048FC27 CRC64;

Query Match 4.38; Score 193; DB 3; Length 451;  
Best Local Similarity 22.88; Pred. No. 0.00013;  
Matches 102; Conservative 63; Mismatches 170; Indels 112; Gaps 15;

QY 284 DDTARELKTPSAERREELRQ-----PYKRLRLNGDWDGTPRARPEREREDQOS 336  
DB 64 EDVGRKHKTKSLKSNDSQKISKKGAPRK-----AHSSSEAGSGSS 109  
QY 337 PNVSIMQMSDMLRWFEASEVAOSNRGRSRRGTSQSDISTLPVPSPDLEVE 396  
DB 110 DE-----SDSSSESSSESDNDSSSDSSSESSSESDSS-----SSDSE-SE 156  
QY 397 TAMEVDTPAEQFLQPSSTSSAQAHSSTSPESHSTPLISPDSEBORQSEASGHHTH 456  
DB 157 SSSSE-----GSQSSSSSSSSSESSSESDNDSS--SSDSESSSEDS-SSS 202  
QY 457 HQSDNNNEKLSPKPGTGTVSTIKLNFDEWSIASSRGIGSHCKSE 516  
DB 203 SSSDSESS-----SSRGSDSSSSSSSESSSESDNDSSSE 246  
QY 517 GQESFVQSSVQPEQSETKAPEESSEDTYKQGVSAENPVENHINITQSDKFTAKP 576  
DB 247 SSSSESDSSSSSDSESSKSDSSSDSSSDSSSDSSSDSSSDSSSDSS-SSS 305  
QY 577 LDSNGERNDLNDRSCGVPEESASSEKAKEPETSDQSTESATNNTNPPQOTAT 636  
DB 306 SSSDSSSSSE-----DGNSTDTTSGEVAQSTNSTSEESTSVKDEDSKIH 355  
QY 637 GPS---AHEE---TSTRDS-----ALQPTDSDDDVYLIPGAR 668  
DB 356 DKSLAKRKHEDDESSSTSTSTKTPTRVGDPSQMDASPALRNSNPFEDDY----- 409  
QY 669 YRAGPDRSAVARIQEFFFFRRRERKE 695  
DB 410 ---GTLANRDLIVTGKGFQEKKKK 433  
RESULT 37  
Q9S255 PRELIMINARY; PRT; 852 AA.  
AC Q9S255;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE Hypothetical 92.0 kDa protein.  
GN F11C18.80 OR A14G3180.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bevan M., Terry N., Ardiles W., Buysshaert C., Dasseville R.,  
 De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Daele H.,  
 Villarroel R., Gielens J., Van Montagu M., Hohelsel J., Mewes H.W.,  
 Mayer K.F.X., Scheller C.;  
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Terry N., Ardiles W., Buysshaert C., Dasseville R., De Clerck R.,  
 De Keyser A., Neyt P., Rouze P., Van Den Daele H., Villarroel R.,  
 Gielens J., Van Montagu M., Mewes H.W., Lemcke K., Mayer K.F.X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL049607; CAB40758.1;  
 DR EMBL: AL161579; CAB79906.1;  
 KW Hypothetical protein.  
 SQ SEQUENCE 852 AA; 91950 MW; AEA128F4D5A0F59D CRC64;  
 Query Match 4.3%; Score 193; DB 10; Length 852;  
 Best Local Similarity 21.9%; Pred. No. 0.0003;  
 Matches 129; Conservative 76; Mismatches 224; Indels 160; Gaps 26;  
 QY 162 EDGTVWFPTKRTKCTKEDCKDDILNCRRAATVAICPIPIYLLAVGSDSSVRIYDR 221  
 DB 365 ESSSVKQADLS-KSDIKETEPALLDSKDVLTSPV-----DSSV----- 405  
 QY 222 RMLGTATGNYAGKRTGTGMARIPSHLNKSCRVTSLCSEGOELVSYSDIYLF 281  
 DB 406 ----TATSS-----ENENKRSV-----DILPSKTSG----- 428  
 QY 282 PKDDTARELKTPSAEERREELRQPVKRLRLGDWSDTGPRAPESEREDGQSPNVL 341  
 DB 429 --DETA-NVSSPMAELELPQSY-----KTIANQKKESSTE--EYKSASI 471  
 QY 342 MQBMSDMLSRWFEASEVAQNSNGRGRSRPGTQSOSDITLTPV--SSPDLEV-- 394  
 DB 472 AT-----EEVSEEPNTSEPQ-VTKKSGKVAASSSKTKPTVPKSKTSKTSKYAKQ 520  
 QY 395 -----SETAMEVDTPAPQFLOPSTSSIMSQA-HSTSSPTESP--HSTPLISSPDSQ 444  
 DB 521 SEKKVGSNDADSTKPKKEKKPGKKAIDESLHTSSGDNKKPAVSSGKLASKSKKKA 580  
 QY 445 ROSVEASGHHHTHOSDNNEKLSPKGTGEPLYSLHSTEGTTSTIKLNTDEMSSIAS 504  
 DB 581 KQTV-----ESPNSMTKRRSLGQGA-----SGESLVGSIKY--WMPDQA 622  
 QY 505 SSRGICSHCKS-----EGQESFVPOSSVOPPEDSKADDESS-----EDVT 548  
 DB 623 YKGVVESYDAKKKHLVYDGDDEILYLNQKWSPLDESELQGEAEADQGEEDAS 682  
 QY 549 KYQEGVSAL--NPENININTQSDKFTAKPLDSSNGERNDLNDSCGVPESSASER 604  
 DB 683 TWGSGGSSKAKATPASKSSKTSQDQKTSKDSKDEASR-----EEEASSEE 730  
 QY 605 AKEPETSQDT---STESATNENNTNPEQFOTEGATGPSAHEETSTDSALQDTDDDDP 661  
 DB 731 EEEEEPPTVTKSGSSSRKDISSYSKSGKASAKKEEBSKATTS-----KSKSGP 785  
 QY 662 VLIIGARYAGQDRRSAAVARIQEEFFRRRERKEEMEDTLNIRPLVK 710  
 DB 786 VKSVPASKTKGKAKSGASSTPA-SKAKESASESESEETPREPEPATK 833

RESULT 38  
 ID 067643 PRELIMINARY; PRT; 985 AA.  
 AC 067643;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE Infectious laryngotracheitis virus US10, US2, protein kinase, UL47,  
 DE glycoprotein G, ORF5, glycoprotein D, glycoprotein I, glycoprotein E,  
 DE ORF9 genes.  
 OS Gallid herpesvirus 1.  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Infectious laryngotracheitis-like viruses.  
 NCBI\_TaxID=10386;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-USDA CHALLENGE STRAIN;  
 RX MEDLINE=97033380; PubMed=8879127;  
 RA Wild M.A., Cook S., Cochran M.;  
 RT "A genomic map of infectious laryngotracheitis virus and the sequence  
 and organization of genes present in the unique short and flanking  
 regions."  
 RL Virus Genes 12:107-116(1996).  
 DR EMBL: U28832; AAC55099.1;  
 SQ SEQUENCE 985 AA; 106520 MW; 89E29D1EB3EB458 CRC64;  
 Query Match 4.3%; Score 192.5; DB 12; Length 985;  
 Best Local Similarity 20.7%; Pred. No. 0.0004;  
 Matches 173; Conservative 101; Mismatches 290; Indels 273; Gaps 41;  
 QY 172 RIKTSCKEKCKDDILNCRRAATVAICPIPIYLLAVGSDSSVRIYDR-RMLGTRATG 230  
 DB 146 RLFTPCDENTLQNEPWWGSKR-----WLGPPSY-----VRQDVAVLTAQYIIGEYSN 195  
 QY 231 NYAGRTGTGVAFET--PSHLAN-----KCRVTSLSYSDGOELVSYSDIYLF 280  
 DB 196 SAAQTGLTSLNMTFFYSKRIYVWTTGGSPSRITIVYSSRNQGVNLNVDGLVYK 255  
 QY 281 DPKDD-----TA-RELKTSABERREELRQPVKRLRLGDWS 317  
 DB 256 TPIDIDRAMINIVANTSPADSGSYLAFTAFREKRLPSAIDH-----RIDWS 302  
 QY 318 DTGPRAPESERERDEQSPNVSIMQMSDMLSRWFEASEVAQNSNGRGRSRPGTQSQ 377  
 DB 303 GTRP--PGTETTFD-----CQKM-----LETPRALGSNVPRODSIR 337  
 QY 378 SDISTLPTVPSSPDLE--VSETAMEVDTPAPQFLOPSTSSIMS--AOAHSTSSPTESP-- 431  
 DB 338 PGATLPPFTDAADPDTGTSPTPTVPEPAITTLIPRSTSDMGFFSTARATGSETLSVPV 397  
 QY 432 -----HSTPL-----LSPDSEQ-----ROSVEASGHHHTHOSDNNEKLSPKG 471  
 DB 398 QETDRLSTTLPPLPLPLPRESENTLFTPTAPGISTEPASAHHTTQTSQAEYVFTQSPS 457  
 QY 472 T-----GEPVLSLHSTEGTTSTIKLNTD--EMSSIASSSRGI-----GS 511  
 DB 458 TESETARSOQSEB-----WFTQPTSTEQALPQTQIAETFAITPQPSAQMFTQTGPA 513  
 QY 512 HCSSEGOESFVP-----QSSVOPPE-GDSETKAPE--ESESEVITYQESVAMENVENH 563  
 DB 514 ETEAPAQDTSTLPEITQSTSTPETARAPASAPBEVFTQSSSTVTEV----- 560  
 QY 564 INITQDKFTAKPLDSSNGERNDLNDSCGVPEESASAKAPEPSTQSTES----- 618  
 DB 561 --FTQPTSTVTPKTLSSSTPEAFTTQSACT--EAFQTSSAEPTMRTQSTETTFQ 616  
 QY 619 -----ATNENNTNPEQFOTEGATGPSAHEETSTDSALQDTDDDDP 650  
 DB 617 APSTVPAKQTQSTEBEVLQSPSTPVPFTRTLGAEPITQPSAAPVYTRSSSTWPE 676  
 QY 651 -LQDTDDSDDDPLTIGAKTRAGPGRRS-----AAVARIQEEFFRRRERKEEMEDTLNIRPLVK 695



Db 677 TAOSTPLASONPT-----SSGTGTHNTPRTYPTVOTPHQKITYENTKLSPTTVVSE 729

Qy 696 MERIDPLNTRRL--VKANYKGRN-----SRNIK-----EANFGANEV-----MS 736

Db 730 FHEMSTAESOTPLDVKYIVEKFSNDGEVATCSTVKSPEYRETN-WKVDLVADVDEIS 788

Qy 737 GSDCGHIF-----IWDHRAEHLMLLEADNHNVNLOPHR---FPPIILASSG 780

Db 789 GNSPACVFNSENKOKOLYRYTDRGTSVQMLCLSTSH-----SPEPCYCLFTSLIAR- 842

Qy 781 IDYDIKIMSPLEERSIFNRKLADVITRNLMEETRNITVTPASPMALASLNHI 837

Db 843 -----EKDIAPELKFTSD--PQTAVCITLPSGVVPRREMLANN 880

RESULT 39

Q9W320 PRELIMINARY; PRT; 2951 AA.

AC Q9W320: 01-MAY-2000 (TREMblrel. 13, Created)

DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)

DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)

DE CG3950 protein.

GN CG3950.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydriodes; Drosophilidae; Drosophila.

OX NCBI\_Taxid=7227;

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RA MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amaratilake P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,

RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borrova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,

RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Matzel B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Modarres C., Morris J., Mostreli A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Munz D.M., Nelson D.T.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleby J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Relbert K., Remington I., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spirdling A.C., Stapleton M., Strong R., Sun E.,

RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,

RA Williams S.M., Woodage T., Morley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveril J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,

RT "The genome sequence of Drosophila melanogaster.";

RL Science 287:2185-2195(2000).

DR EMBL; AE003438; AAF46171.1; -

DR FlyBase; Fggn0029875; CG3950.

SO SEQUENCE 2951 AA; 332269 MW; 85CC0A8BDA8AEB CRC64;

Query Match 4.2%; Score 191.5; DB 5; Length 2951;

Best Local Similarity 21.5%; Pred. No. 0.002;

Matches 108; Conservative 72; Mismatches 181; Indels 141; Gaps 22;

Qy 273 SSDYIYLPDRKD-TAEELK-----TPSAERREELRPPYRLRLRGWSTPG-PRAPES 327

Db 455 TSEIVEYEDDRDPTSHVQAKPPSPYAPPSHSPROSPAK-----DFSTHGFPSPRNK 508

Qy 328 EREHDEGQSPNVS-----LMQRM-----DMLRWFEASVQAQNR-- 364

Db 509 ATQETPEQRCIGIAEEVYVYRSEKSRQYKQOTSQRTIEYEGDDYQEPQRFQKREA 568

Qy 365 ---GRGRSRPRGTSQSDIST--LPTV-----PSSPDLVSEETAEVDTAPDFLPQ 411

Db 569 PTPSWEDPQRTQROVEEDFTHGFPVRYTTTSTRPDQGEVLHTKTVSRNQSARKT 628

Qy 412 STSTMSAQASHSTSPESHSTPILSSPDSQASQASGHHHOSDNNEKLSKPG 471

Db 629 NTERLIETQVEHPNAPS--HSTPRSGSPRSQPRDYAS-----SPRGP 670

Qy 472 TGEPLVSLHSTEGTSTIKLNTDE-----WSTIASS 506

Db 671 AGRPDSQSRPNTGSGTIT--KTTTSEPLIRROLQKREVDAAHRAAASLRSSPADST 728

Qy 507 RGIGSHCKSEGOSEFVQSSV-----OPRGDETAPESSE---DYTKYQEG- 553

Db 729 TSYGSHHQ-----PRSSVSSNRTFRREMRGSHDSQAPSSSSISSTYTRHHTGG 780

Qy 554 -----VSAENPVENHINIQSDKFT--AKPLDSNGERNLNL--DRSCGVPESSASS 602

Db 781 NYSMTIKTKITTKRNHKKKAIEPATDTSQPIRKVLSANEAQVVEACVRRQYOLG 840

Qy 603 EKAKEPETSQTSATESATNEN-----NTNPEQOTEATGQSAHEESTDSAL----- 651

Db 841 NGENEPETPSSSGTPAKKRRHQMRPPDEPFPOLRRSSKSPSYPRQYQKHTTPEGRRVS 900

Qy 652 QDPTDDSDDDPVL-----PGA 667

Db 901 QDREISIDELLILETSGARGS 922

RESULT 40

Q9R0L0 PRELIMINARY; PRT; 1508 AA.

AC Q9R0L0: 01-MAY-2000 (TREMblrel. 13, Created)

DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)

DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)

DE JNK-binding protein JNKB1.

GN JNKB1 OR JNKB1.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_Taxid=10090;

RP SEQUENCE FROM N.A.

RC TISSUE=BRN;

RA MEDLINE=99402957; PubMed=10471813;

RA Koyano S., Ito M., Takamatsu N., Shiba T., Yamamoto K., Yoshioaka K.,

RT "A novel Jun N-terminal kinase (JNK)-binding protein that enhances the

RT activation of JNK by MEK kinase 1 and TGF-beta-activated kinase 1.";

RL FEBS Lett. 457:385-388(1999).

CC -1- SIMILARITY: CONTAINS 12 WD REPEATS (THP-ASP DOMAINS).

DR EMBL; AB029482; BAA85449.1; -

DR MGI; MGI:1347004; Jnkb1.

DR InterPro; IPR002114; HPr\_Serp\_site.

DR InterPro; IPR001680; WD40.

DR Pfam; PF00400; WD40; 12.

DR SMART; SM00320; WD40; 11.

DR PROSITE; PS00589; PTS\_HPR\_SER; UNKNOWN\_1.

DR PROSITE; PS50082; WD\_REPEATS\_2; 1.



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 18, 2003, 15:35:56 ; Search time 17 Seconds  
(without alignments)  
2331.717 Million cell updates/sec

Title: US-09-781-693a-2

Perfect score: 4527

Sequence: 1 MSRGSYPHLLWDVRRKRLG.....RLEGDRSGQENEDDEE 860

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 199416 seqs, 46092074 residues

Total number of hits satisfying chosen parameters: 199416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing First 120 summaries

Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubaa/US08\_NEW\_PUB pep.\*
- 2: /cgn2\_6/ptodata/2/pubaa/PCT\_NEW\_PUB pep.\*
- 3: /cgn2\_6/ptodata/2/pubaa/US06\_PUBCOMB pep.\*
- 4: /cgn2\_6/ptodata/2/pubaa/US07\_NEW\_PUB pep.\*
- 5: /cgn2\_6/ptodata/2/pubaa/US07\_PUBCOMB pep.\*
- 6: /cgn2\_6/ptodata/2/pubaa/PCTUS\_PUBCOMB pep.\*
- 7: /cgn2\_6/ptodata/2/pubaa/US08\_PUBCOMB pep.\*
- 8: /cgn2\_6/ptodata/2/pubaa/US09\_PUBCOMB pep.\*
- 9: /cgn2\_6/ptodata/2/pubaa/US09\_NEW\_PUB pep.\*
- 10: /cgn2\_6/ptodata/2/pubaa/US10\_PUBCOMB pep.\*
- 11: /cgn2\_6/ptodata/2/pubaa/US10\_NEW\_PUB pep.\*
- 12: /cgn2\_6/ptodata/2/pubaa/US10\_PUBCOMB pep.\*
- 13: /cgn2\_6/ptodata/2/pubaa/US60\_NEW\_PUB pep.\*
- 14: /cgn2\_6/ptodata/2/pubaa/US60\_PUBCOMB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	569	12.6	110	10	US-09-864-761-35853
2	437	9.7	86	10	US-09-864-761-33782
3	338	7.5	62	10	US-09-864-761-33783
4	273	6.0	214	10	US-09-925-300-1585
5	190.5	4.2	1367	10	US-09-801-368-108
6	189	4.2	1142	9	US-10-085-108-7
7	189	4.2	1142	10	US-09-899-651-2
8	188	4.2	1702	9	US-09-839-996-5
9	185.5	4.1	571	10	US-09-765-272-4
10	183	4.0	2344	10	US-09-815-242-12713
11	180.5	4.0	1175	10	US-09-771-161A-224
12	180.5	4.0	1175	10	US-09-771-161A-225
13	180.5	4.0	1175	10	US-09-771-161A-226
14	179.5	4.0	1501	10	US-09-924-154-17
15	178	3.9	1325	10	US-09-864-761-35612
16	175.5	3.9	1848	9	US-09-839-996-6
17	174	3.8	2368	10	US-09-815-242-5635
18	174	3.8	2368	10	US-09-815-242-12389
19	170	3.8	684	10	US-09-925-302-555

20	169	3.7	1421	10	US-09-924-154-13	Sequence 13, Appl
21	167.5	3.7	1031	10	US-09-815-242-10932	Sequence 10932, A
22	166	3.7	560	10	US-09-815-242-13057	Sequence 13057, A
23	165.5	3.7	1435	9	US-10-153-273-4	Sequence 4, Appli
24	164.5	3.6	957	9	US-10-025-380-1065	Sequence 1065, Ap
25	164.5	3.6	957	10	US-09-922-217-1065	Sequence 1065, Ap
26	164.5	3.6	957	10	US-09-833-263-1065	Sequence 1065, Ap
27	164.5	3.6	2478	10	US-09-815-242-5816	Sequence 5816, Ap
28	164.5	3.6	2478	10	US-09-815-242-12967	Sequence 12967, A
29	164	3.6	278	10	US-09-765-272-94	Sequence 94, Appl
30	160.5	3.5	1356	9	US-10-077-111-10	Sequence 10, Appl
31	160	3.5	502	10	US-09-815-242-5904	Sequence 5904, Ap
32	156.5	3.5	713	10	US-09-801-368-498	Sequence 498, App
33	155.5	3.4	505	9	US-09-820-843A-12	Sequence 12, Appl
34	155.5	3.4	505	10	US-09-815-242-11317	Sequence 11317, A
35	155.5	3.4	617	10	US-09-864-761-36182	Sequence 36182, A
36	154.5	3.4	729	9	US-10-145-396-11	Sequence 11, Appl
37	152.5	3.4	980	9	US-10-108-605-195	Sequence 195, App
38	152.5	3.4	1781	10	US-09-738-877-3	Sequence 3, Appli
39	151.5	3.3	1139	9	US-09-820-843A-15	Sequence 15, Appl
40	151	3.3	1075	10	US-09-801-368-110	Sequence 110, App
41	149.5	3.3	748	10	US-09-864-761-43244	Sequence 43244, A
42	149.5	3.3	1322	10	US-09-801-368-114	Sequence 114, App
43	149	3.3	473	10	US-09-864-761-38321	Sequence 38321, A
44	149	3.3	1162	10	US-09-745-008-34	Sequence 34, Appl
45	148.5	3.3	897	10	US-09-815-242-12769	Sequence 12769, A
46	148.5	3.3	1113	10	US-09-815-242-5836	Sequence 5836, Ap
47	148.5	3.3	1596	9	US-09-902-432-4	Sequence 4, Appli
48	148	3.3	1125	9	US-09-974-298-114	Sequence 114, App
49	148	3.3	1163	9	US-09-932-257A-3	Sequence 3, Appli
50	147.5	3.3	511	10	US-09-864-761-34590	Sequence 34590, A
51	147.5	3.3	2843	9	US-09-987-482-1	Sequence 1, Appli
52	147	3.2	1360	10	US-09-871-916-2	Sequence 2, Appli
53	146	3.2	1537	10	US-09-801-368-104	Sequence 104, App
54	145.5	3.2	596	9	US-10-063-547-100	Sequence 100, App
55	145.5	3.2	596	9	US-10-174-590-310	Sequence 310, App
56	145.5	3.2	596	9	US-10-176-758-310	Sequence 310, App
57	145.5	3.2	596	9	US-10-063-616-100	Sequence 100, App
58	145.5	3.2	596	9	US-10-175-737-310	Sequence 310, App
59	145.5	3.2	596	9	US-10-073-502-100	Sequence 100, App
60	145.5	3.2	596	9	US-10-173-706-310	Sequence 310, App
61	145.5	3.2	596	9	US-10-175-738-310	Sequence 310, App
62	145.5	3.2	596	9	US-10-175-738-310	Sequence 310, App
63	145.5	3.2	596	9	US-10-176-482-310	Sequence 310, App
64	145.5	3.2	596	9	US-10-176-913-310	Sequence 310, App
65	145.5	3.2	596	9	US-10-176-913-310	Sequence 310, App
66	145.5	3.2	596	9	US-10-180-552-310	Sequence 310, App
67	145.5	3.2	596	9	US-10-180-552-310	Sequence 310, App
68	145.5	3.2	596	9	US-10-173-700-310	Sequence 310, App
69	145.5	3.2	596	9	US-10-174-572-310	Sequence 310, App
70	145.5	3.2	596	9	US-10-174-572-310	Sequence 310, App
71	145.5	3.2	596	9	US-10-174-582-310	Sequence 310, App
72	145.5	3.2	596	9	US-10-174-582-310	Sequence 310, App
73	145.5	3.2	596	9	US-10-175-739-310	Sequence 310, App
74	145.5	3.2	596	9	US-10-175-740-310	Sequence 310, App
75	145.5	3.2	596	9	US-10-175-743-310	Sequence 310, App
76	145.5	3.2	596	9	US-10-176-488-310	Sequence 310, App
77	145.5	3.2	596	9	US-10-176-493-310	Sequence 310, App
78	145.5	3.2	596	9	US-10-176-747-310	Sequence 310, App
79	145.5	3.2	596	9	US-10-176-750-310	Sequence 310, App
80	145.5	3.2	596	9	US-10-176-985-310	Sequence 310, App
81	145.5	3.2	596	9	US-10-176-985-310	Sequence 310, App
82	145.5	3.2	596	9	US-10-176-991-310	Sequence 310, App
83	145.5	3.2	596	9	US-10-176-992-310	Sequence 310, App
84	145.5	3.2	596	9	US-10-176-993-310	Sequence 310, App
85	145.5	3.2	596	9	US-10-184-658-310	Sequence 310, App
86	145.5	3.2	596	9	US-10-173-695-310	Sequence 310, App
87	145.5	3.2	596	9	US-10-173-697-310	Sequence 310, App
88	145.5	3.2	596	9	US-10-173-705-310	Sequence 310, App
89	145.5	3.2	596	9	US-10-174-576-310	Sequence 310, App
90	145.5	3.2	596	9	US-10-174-585-310	Sequence 310, App
91	145.5	3.2	596	9	US-10-174-586-310	Sequence 310, App
92	145.5	3.2	596	9	US-10-175-747-310	Sequence 310, App

93 145.5 3.2 596 9 US-10-176-481-310 Sequence 310, App  
94 145.5 3.2 596 9 US-10-176-485-310 Sequence 310, App  
95 145.5 3.2 596 9 US-10-176-487-310 Sequence 310, App  
96 145.5 3.2 596 9 US-10-176-493-310 Sequence 310, App  
97 145.5 3.2 596 9 US-10-176-756-310 Sequence 310, App  
98 145.5 3.2 596 9 US-10-176-911-310 Sequence 310, App  
99 145.5 3.2 596 9 US-10-176-919-310 Sequence 310, App  
100 145.5 3.2 596 9 US-10-176-925-310 Sequence 310, App  
101 145.5 3.2 596 9 US-10-176-978-310 Sequence 310, App  
102 145.5 3.2 596 9 US-10-179-510-310 Sequence 310, App  
103 145.5 3.2 596 9 US-10-180-543-310 Sequence 310, App  
104 145.5 3.2 596 9 US-10-180-544-310 Sequence 310, App  
105 145.5 3.2 596 9 US-10-180-546-310 Sequence 310, App  
106 145.5 3.2 596 9 US-10-180-547-310 Sequence 310, App  
107 145.5 3.2 596 9 US-10-180-549-310 Sequence 310, App  
108 145.5 3.2 596 9 US-10-180-555-310 Sequence 310, App  
109 145.5 3.2 596 9 US-10-180-559-310 Sequence 310, App  
110 145.5 3.2 596 9 US-10-181-000-310 Sequence 310, App  
111 145.5 3.2 596 9 US-10-183-010-310 Sequence 310, App  
112 145.5 3.2 596 9 US-10-183-012-310 Sequence 310, App  
113 145.5 3.2 596 9 US-10-184-614-310 Sequence 310, App  
114 145.5 3.2 596 9 US-10-184-623-310 Sequence 310, App  
115 145.5 3.2 596 9 US-10-184-635-310 Sequence 310, App  
116 145.5 3.2 596 9 US-10-184-637-310 Sequence 310, App  
117 145.5 3.2 596 9 US-10-184-646-310 Sequence 310, App  
118 145.5 3.2 596 9 US-10-184-647-310 Sequence 310, App  
119 145.5 3.2 596 9 US-10-184-652-310 Sequence 310, App  
120 145.5 3.2 596 9 US-10-187-594-310 Sequence 310, App

## ALIGNMENTS

RESULT 1  
US-09-864-761-35853  
; Sequence 35853, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aecomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ IDS: 49117  
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 35853  
; LENGTH: 110  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AL033531.7  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.8  
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.3  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.8  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.99  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.4  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.7  
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.9  
; OTHER INFORMATION: EST\_HUMAN HIT: A1927915.1, EVALUATE 8.00e-35  
US-09-864-761-35853

Query Match 12.6%; Score 569; DB 10; Length 110;  
Best Local Similarity 100.0%; Pred. No. 7.2e-28;  
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 568 QSDKFTAKPLDLSNGERNDLNRSCGVPESASSEKAKPEPDSQSTESATNNTNP 627  
Db 1 QSDKFTAKPLDLSNGERNDLNRSCGVPESASSEKAKPEPDSQSTESATNNTNP 60  
QY 628 EPQFQTEATGPSAHEETSTRDSALQDITDDSDDDPVLIPGARYRAGPGR 676  
Db 61 EPQFQTEATGPSAHEETSTRDSALQDITDDSDDDPVLIPGARYRAGPGR 109

## RESULT 2

US-09-864-761-33782  
; Sequence 33782, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aecomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
SEQ ID NO 33782  
LENGTH: 86  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AL031287.3  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.6  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.95  
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.3  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.4  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.81  
OTHER INFORMATION: EST HUMAN HIT: 238367.1, EVALUE 5.00e-25  
OTHER INFORMATION: SWISSPROT HIT: Q64752, EVALUE 3.10e+00  
OTHER INFORMATION: EST HUMAN HIT: BE867206.1, EVALUE 5.00e-29  
US-09-864-761-33782

Query Match 9.7%; Score 437; DB 10; Length 86;  
Best Local Similarity 100.0%; Pred. No. 5e-20;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 374 GTSQSDISTLPTVPSPDLEVSSTAMEVDTPAQFLQPTSTSTMSQAQHTSSPTSPHS 433  
Db 1 GTSQSDISTLPTVPSPDLEVSSTAMEVDTPAQFLQPTSTSTMSQAQHTSSPTSPHS 60  
QY 434 TPLLSPDSEQRQSVASGHHTHQS 459  
Db 61 TPLLSPDSEQRQSVASGHHTHQS 86

RESULT 3  
US-09-864-761-33783  
Sequence 33783, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aecmca-x-1  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
SEQ ID NO 33783  
LENGTH: 62  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AL031287.3  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.2  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2  
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.4  
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.4  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.3  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2  
OTHER INFORMATION: EST HUMAN HIT: BE019684.1, EVALUE 5.00e-32  
OTHER INFORMATION: EST HUMAN HIT: BE180609.1, EVALUE 8.00e-32  
OTHER INFORMATION: SWISSPROT HIT: P34311, EVALUE 9.90e+00  
US-09-864-761-33783

Query Match 7.5%; Score 338; DB 10; Length 62;  
Best Local Similarity 100.0%; Pred. No. 3.1e-14;  
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 85 VLTIRSGHRANIFSAKFLPCTNDKQIVSCGSGDVIFYTNVEQDAETNRCQFTCHYGT 144  
Db 1 VLTIRSGHRANIFSAKFLPCTNDKQIVSCGSGDVIFYTNVEQDAETNRCQFTCHYGT 60  
QY 145 YE 146  
Db 61 YE 62

RESULT 4  
US-09-925-300-1585  
Sequence 1585, Application US/09925300  
Patent No. US20020151681A1  
GENERAL INFORMATION:

APPLICANT: Craig Rosen,  
APPLICANT: Steve Ruben  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA101  
CURRENT APPLICATION NUMBER: US/09/925,300  
CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05988  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 1890  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1585  
LENGTH: 214  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (1)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (2)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (3)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (12)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-300-1585

Query Match 6.0%; Score 273; DB 10; Length 214;  
Best Local Similarity 36.2%; Pred. No. 1.4e-09;  
Matches 59; Conservative 33; Mismatches 53; Indels 18; Gaps 5;

QY 713 YGHRNRTMIKEANFWG--ANFVMSGDCGHIFWDRHTAHLMLLEAD-NHVVNCLQP 769  
DB 45 YGHRNAT-VKGVNFGPKSEFVYVSGDCGHIFLWKSQOIQFMGDKGVVNCLEP 103  
QY 770 HPFDPTIASSGDYDIKTIWSPLEESRIFNRKLADEVITRNELMLETETNTITVPASEMLR 829  
DB 104 HPHLPVATSGLDHVDKIWAPTAESTELTGLKOVIKKKNKREDEDSLHQTDLFDSHLW 163  
QY 830 MIAASLNIHADR-----LEGDRSGSGQENENEDEE 860  
DB 164 FL-MHLRQRHRRWRPVGATDADSDESPSSSDTSDEE 204

RESULT 5  
US-09-801-368-108  
Sequence 108, Application US/09801368  
Patent No. US20020128250A1  
GENERAL INFORMATION:

APPLICANT: Busby, Robert  
APPLICANT: Cali, Brian  
APPLICANT: Hecht, Peter  
APPLICANT: Holtzman, Doug  
APPLICANT: Madden, Kevin  
APPLICANT: Maxon, Mary  
APPLICANT: Milne, Todd  
APPLICANT: No. US20020128250A1man, Thea  
APPLICANT: Royer, John  
APPLICANT: Salama, Sofie  
APPLICANT: Sherman, Amir  
APPLICANT: Silva, Jeff  
APPLICANT: Summers, Eric  
TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi  
FILE REFERENCE: 109272.147  
CURRENT APPLICATION NUMBER: US/09/801,368  
CURRENT FILING DATE: 2001-03-07  
PRIOR APPLICATION NUMBER: US 09/487,558  
PRIOR FILING DATE: 2000-01-19  
PRIOR APPLICATION NUMBER: US 60/160,587

PRIOR FILING DATE: 1999-10-20  
NUMBER OF SEQ ID NOS: 440  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 108  
LENGTH: 1367  
TYPE: PRT  
ORGANISM: Saccharomyces cerevisiae  
US-09-801-368-108

Query Match 4.2%; Score 190.5; DB 10; Length 1367;  
Best Local Similarity 22.3%; Pred. No. 0.0015;

Matches 124; Conservative 64; Mismatches 270; Indels 99; Gaps 18;

QY 151 PNDPYFLSCGEDGTVMFDTTRIKTSCTEDCKDDLLINCRRAATVAICPIPIYLAAG 210  
DB 272 PTPPTTSTCKEPTPPHDT---TPCTKK--KTTSKCTCKTKTT-----VPVTPSSST 321  
QY 211 CSDSSVRIYDRMLGTRATGNYAGRTTGVARFIP-----SHLNKSKRVTSILCYSEDQ 266  
DB 322 TESSAPVPTPSSSTTESSAPVTSSTTESSAPVPTPSSSTTESSAPVTSSTTESSA 381  
QY 267 EILVSYSSDYIYLFDPKDDTARELKTPSAERREELRQPPVKRLRLRGDWSDFGPAREE 326  
DB 382 PVTSSTT-----ESSAPVPTPSS-----STTESSAPV 410  
QY 327 SERERDEQSPVSLMQMSDMLSRFEEASEVAQSNRGRSRP-RGGTSQSDISTLPT 385  
DB 411 TSSTTESSAP-----VTSTTESSAPVTSSTTESSAPVTSSTTESSAPVPT 460  
QY 386 VPSSPDLEVESE---TAMEVDTPAEQFLQSTSTMSAQAHSTSPTES---PHSTPLLS 439  
DB 461 -PSSSTTESSAPVTSSTTESSAPVPTPSSSTTESSAPVTSSTTESSAPVPTPSSST 519  
QY 440 PDSEQRQSVASGHHTHQH-----DNNNEKLS---PKPGTGEVPLSLHYSTEGTTSTIK 492  
DB 520 TESSAPAPTPSSSTTESSAPVTSSTTESSAPVPTPSSSTTESSSTPVTSTTESSA 579  
QY 493 LNTDEWSSIASRRGIGSHCKSEGGQESF---VPOSSVQPPGEDSKAPESSSEDVTK 549  
DB 580 PVPTPSSSTTESSAPVPTPSSSTTESSAPAPTPPSSSTTESSAPVTSSTTESSAPV 639  
QY 550 YQEGVSAEN---PVENHINITQSDKFTAKPLDINSRNDLNLDRSCGVPEESASSEKAK 606  
DB 640 TPSSSTTESSAPVPTPSSSTTESSAPVPTPSS-----TTESSAPVTSSTTESS 692  
QY 607 EPETSDQTSESA-----TNENNTNPEPQCTEAT-----GPSAHEETSTDSA 650  
DB 693 APVTSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPV 752  
QY 651 LQDTDDSDDDPVLIPGA 667  
DB 753 TSSTTESSAPVPTPSS 769

RESULT 6

US-10-085-108-7  
Sequence 7, Application US/10085108  
Patent No. US20020176865A1

GENERAL INFORMATION:  
APPLICANT: Lucas, Sophie; BOON-FALLEUR, Thierry  
TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES CODING FOR  
TUMOR REJECTION ANTIGEN PRECURSORS OF MEMBERS OF THE MAG

NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESS:  
ADDRESSSEE: Fulbright & Jaworski L.L.P.  
STREET: 666 Fifth Avenue  
CITY: New York City  
STATE: New York  
COUNTRY: USA  
ZIP: 10103

COMPUTER READABLE FORM:

\_\_\_\_\_

100





```
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 571 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-765-272-4

Query Match
Best Local Similarity 4.1%; Score 185.5; DB 10; Length 571;
Matches 99; Conservative 80; Mismatches 198; Indels 137; Gaps 17;

QY 212 SDSSVRIYDRMLGRATGNYAGRTTGMVARFIPSHLNKSCRVTSCLYSEDGOFILYS 271
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 110 SDSQQLAEHKNLETKKEKISPKETKGVNTLNPQDEVLSGQLNKPPELYREETMETKD 169

QY 272 YSS-----DYIYFD-PKDDTARELKTPSAERRELRQP 305
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 170 FOEIQENPDLAGTVRVKQEGKLGKKEIVRIFSVNKEEVSREIVSTST-----TA 221

QY 306 PVKRLRLRGDMSDTGPRAPSERERDGEOS-----PNVSLMQMSDMLSRFEEASEVA 360
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 222 PSPRIVEKGTQVTKQEPGEVGHKDVQSGAIVEPAIQ-----PELPEAV 268

QY 361 QSNRGGRSRPRGGTSQSDISTPLTPV-----PSSPDLEYSETA-----S 790
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 269 VSDKGEPEVQ-----TLPEAVTDKETEVEQPESDTVVSDKGEPEQVAPLPEY 318

QY 399 -----MEVDTPAEQFLOPSSISMSAQAHSTSTSPHSTPLSSPDSEORQSVASG 452
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 319 KNIEQVAPETVEKTKQGPKEVEV-----PVKPTPEVPNPNEGTTEGTIOEAE 371

QY 453 HHTH--HOSDNNNEKSPKPGTGPVLSLHYSTEGTTSTIKLNTDEWSSIASRRGIG 510
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 372 NPVQPAESTINSEKVP-----DTSKNTGEVSSNPSDSTTSVGENKKEPH 418

QY 511 SHCKSEGESEFVPOSSVQPPGDSSETKAPESSEDEVTKYQGVSAENPVENHINITSD 570
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 419 NDSKNENSEKT-VEEVVPNPNEGTVEGTSNOETKPVQPAER-----TQTN 463

QY 571 KFTAKPLDSNGERDNLDRSCGVPEESASSEK---AKEPETSQTSATNENNTNP 627
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 464 --SGIANENTGEVSNKPSDKPPV-EESNQPEKNGTKPKPENSNTSEN-----QTEP 516

QY 628 EPQFQTEATGPSAHEETSTRSALQDTSDDDP 661
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 517 EPSNGNSTEDVSTESNTSNGNEEKQENELDP 550

RESULT 10
US-09-815-242-12713
; Sequence 12713, Application us/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
```

```
FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12713
; LENGTH: 2344
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12713

Query Match
Best Local Similarity 4.0%; Score 183; DB 10; Length 2344;
Matches 135; Conservative 85; Mismatches 237; Indels 178; Gaps 25;

QY 51 DCVNTICWNDGEYILSGSDDTKLVISPNYSRKVTTTIRSGHRANIFSAKFLPCTNDKQ 110
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 710 DSNNTISGPTN-----IGTITIVTSDAGNKTTTTFKYEIVTRNSMS-----D 755

QY 111 IVSCSGDGVIFTVNBQDAETNRQCFCHYGTTYEIMTVPNPDYTFLSCGEDGTVRWF 170
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 756 SVTSGS-----TQSQSVSTSKADSQSASTSGSIMT-----S 790

QY 171 TRIKTSCTKEDKDDILNCRRAATSAVAICPPPIPYLVAGCSDSVRIYDRMLGRATG 230
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 791 TGASTS-----KTSVSL-----SDS-----VSASKSLSTSEN 819

QY 231 NYAGRTTGMV-ARFTPSHLNKNKSCRVTSCLYSEDGQEIILVSYSDYIYLPKDDTARE 289
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 820 SVSSTSTSLVNSQSVSSMSGVSSTSL-----SDFI-----853

QY 290 LKTPSAEERRELRQPPVKRLRLRGDWD-----TGPRAPSERERDGEQSPNVSLMQ 343
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 854 -SNSSTKESVSSTSDSLSTSLSDSVSMSTSGSLSKSQS---LSTSTSDSASTSQ 909

QY 344 RMSDMLSRFEEASEVAQNRRGRSRPRGGTSQSDISTPLTPV-----PSSPDLEYSETA 398
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 910 SVSDSTSNISSTSELSSES-----GSTSES-ISISNSISNSVSASTSKLESQSTS 958

QY 399 MEVDT-----PAEQFLOPSTSTMS-----AQAHSTSSPTESPHSTPLSSPD-----441
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 959 ISLSTSDSKSMSTSELSSTSTSDSVSGSLSVAGSQSVSTSTSDSMSEMISMSMSTS 1018

QY 442 ----SEQQSVASGHHTHHQDNNNEKLSKPKPGTGE---PVLSLHYSTEGTTSTIKLN 494
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1019 GSLAASDQSKMSVSSMSSTSGSTSELSSTSDSTSDSKSLSTSGSGSTSTS---1074

QY 495 FTDEWSSSIASSRGIGCHCKSEGQEEFVQSSVQPPGDSSETKAPESSEDEVTKYQEV 554
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1075 -TSTSSVRMSESQSTSGSMSTSQSDSTSTSTSTSDSTSDSKS-ASTASSESIS---QSV 1129

QY 555 SAENPVENHINITQDKFTAKPLDSNSGERDNLNDRSCGVPEESASSEKAKEPEPESDT 614
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1130 STS-----TSGSVSTSTSLSTSNSTSTSM-----SDSTSLSTSESDSTSDST 1173

QY 615 STESATNENNTNPQFQTEATGPSAHEETSTRDS 649
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1174 STSDSISEAISG-----SESTISLSESNSTSDS 1202
```

## RESULT 11

US-09-771-161A-224  
; Sequence 224, Application US/09771161A  
; Patent No. US20020110811A1  
; GENERAL INFORMATION:  
; APPLICANT: LEVINE, et al.  
; TITLE OF INVENTION: VARIANTS OF PROTEIN KINASES  
; FILE REFERENCE: 802620-2005.1  
; CURRENT APPLICATION NUMBER: US/09/771,161A  
; CURRENT FILING DATE: 2001-01-26  
; PRIOR APPLICATION NUMBER: 09/724,676  
; PRIOR FILING DATE: 2000-11-28  
; PRIOR APPLICATION NUMBER: 136776  
; PRIOR FILING DATE: 2000-06-15  
; PRIOR APPLICATION NUMBER: 135619  
; PRIOR FILING DATE: 2000-04-12  
; NUMBER OF SEQ ID NOS: 273  
; SOFTWARE: Patent version 3.0  
; SEQ ID NO 224  
; LENGTH: 1175  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-771-161A-224

Query Match 4.0%; Score 180.5; DB 10; Length 1175;  
Best Local Similarity 20.7%; Pred. No. 0.0049;  
Matches 133; Conservative 86; Mismatches 260; Indels 163; Gaps 28;

```
QY 283 KDDTARELATPSAEERREELRQPPVKRLRLGDMSTGPRARPESEERERDGEQSPNVSLM 342
Db 373 EQQRREARQOQREQRREKRR-L-EELERRRKEEERAAEKKRRV--- 424

QY 343 QRMSDMLSRWFEEAS---EVAQSN-----RGRGRSRPRGG 374
Db 425 EREQEYIRQLLEEQRHLEVLQQLLQEQAMLLHRRHPHQSHSQPPPPQERSKPSFH 484

QY 375 TSQSDISTLP-----TVPSDDLEVSETAMEVDTP-----AQOFLQPSSTSM 417
Db 485 APEPKAHYEPADRAREVEDFRKTNHSSPEAQSKQGRVLEPPVPSRSEFSNGNSESVH 544

QY 418 SA--QAHSSTSPTEPHSTPLSSPDSEQRQSVQSGHHTHQSDNNN----- 463
Db 545 PALQRAEPQVPTVTRSRPVLRRDS----PLOGSQQNSQAGQNRNSTSSIEPRLWLR 600

QY 464 -EKLSPKPGTGEVLSHYSTEGTTTSTIKLNFDEWSSIASSSRGIG-----SHCKSE 516
Db 601 VEKLVPRPGSG-----SSGSS-----NSGSPGSHPGSGSGGERFVRSSSKSE 646

QY 517 GQESFVPOSSVQPPGDSQETKAP--EESSEDYTKYQEGVSAENPVENHINITQSKFTA 574
Db 647 GSPSQRL-ENAVKKPEDKKEVFRPLKPAGEVDLTALAKELRAVEDYRPPHKVTD---YSS 702

QY 575 KPLDSNGERNLNL-----DRSCGVPEE--SASSEKAKEPETSDQSTESATNENNTNP 627
Db 703 SSESSTTDEDDDVQEGADESTSGPDTRAASSLNLSNGETE---SVKTMIVHDDVES 759

QY 628 EPQFQTEATGPSAHEETSTRDSALQTDSDDD-----DPVLI----- 664
Db 760 EPAMTPSKEGTLIVRQTQASSTLQKHSSSTPTFPIDPRLQLISPSGGTTVTTSVVGFS 819

QY 665 PGARYAGPGD--RRSAVARIQEFFRRRERKEMEELDTLNIRPLVKVYKGRHSRTM 722
Db 820 DGMRPEAIRQDPTRKGSVNVNPTNTR-----POSTDPEIRK-----YKRFNSEIL 866

QY 723 IREANFWANFVMSGDCGHIFWDRHTAEHLMLLE--ADNHVNCVLOPHPPDPILASSG 780
Db 867 C--AALWGNLVV-GTESG-----IMLLDRSGGQKVPYPLINRRRFOQMDVLEG 911

QY 781 IDYDIKWSPLEESRIFNRKLADEVITRNELMLEETNTITV 822
Db 912 LNVLTISGKKDLRYVYLSWLRNKLHNDPEVEKKQGTWTV 953
```

## RESULT 12

US-09-771-161A-225  
; Sequence 225, Application US/09771161A  
; Patent No. US20020110811A1  
; GENERAL INFORMATION:  
; APPLICANT: LEVINE, et al.  
; TITLE OF INVENTION: VARIANTS OF PROTEIN KINASES  
; FILE REFERENCE: 802620-2005.1  
; CURRENT APPLICATION NUMBER: US/09/771,161A  
; CURRENT FILING DATE: 2001-01-26  
; PRIOR APPLICATION NUMBER: 09/724,676  
; PRIOR FILING DATE: 2000-11-28  
; PRIOR APPLICATION NUMBER: 136776  
; PRIOR FILING DATE: 2000-06-15  
; PRIOR APPLICATION NUMBER: 135619  
; PRIOR FILING DATE: 2000-04-12  
; NUMBER OF SEQ ID NOS: 273  
; SOFTWARE: Patent version 3.0  
; SEQ ID NO 225  
; LENGTH: 1175  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-771-161A-225

Query Match 4.0%; Score 180.5; DB 10; Length 1175;  
Best Local Similarity 20.7%; Pred. No. 0.0049;  
Matches 133; Conservative 86; Mismatches 260; Indels 163; Gaps 28;

```
QY 283 KDDTARELATPSAEERREELRQPPVKRLRLGDMSTGPRARPESEERERDGEQSPNVSLM 342
Db 373 EQQRREARQOQREQRREKRR-L-EELERRRKEEERAAEKKRRV--- 424

QY 343 QRMSDMLSRWFEEAS---EVAQSN-----RGRGRSRPRGG 374
Db 425 EREQEYIRQLLEEQRHLEVLQQLLQEQAMLLHRRHPHQSHSQPPPPQERSKPSFH 484

QY 375 TSQSDISTLP-----TVPSDDLEVSETAMEVDTP-----AQOFLQPSSTSM 417
Db 485 APEPKAHYEPADRAREVEDFRKTNHSSPEAQSKQGRVLEPPVPSRSEFSNGNSESVH 544

QY 418 SA--QAHSSTSPTEPHSTPLSSPDSEQRQSVQSGHHTHQSDNNN----- 463
Db 545 PALQRAEPQVPTVTRSRPVLRRDS----PLOGSQQNSQAGQNRNSTSSIEPRLWLR 600

QY 464 -EKLSPKPGTGEVLSHYSTEGTTTSTIKLNFDEWSSIASSSRGIG-----SHCKSE 516
Db 601 VEKLVPRPGSG-----SSGSS-----NSGSPGSHPGSGSGGERFVRSSSKSE 646

QY 517 GQESFVPOSSVQPPGDSQETKAP--EESSEDYTKYQEGVSAENPVENHINITQSKFTA 574
Db 647 GSPSQRL-ENAVKKPEDKKEVFRPLKPAGEVDLTALAKELRAVEDYRPPHKVTD---YSS 702

QY 575 KPLDSNGERNLNL-----DRSCGVPEE--SASSEKAKEPETSDQSTESATNENNTNP 627
Db 703 SSESSTTDEDDDVQEGADESTSGPDTRAASSLNLSNGETE---SVKTMIVHDDVES 759

QY 628 EPQFQTEATGPSAHEETSTRDSALQTDSDDD-----DPVLI----- 664
Db 760 EPAMTPSKEGTLIVRQTQASSTLQKHSSSTPTFPIDPRLQLISPSGGTTVTTSVVGFS 819

QY 665 PGARYAGPGD--RRSAVARIQEFFRRRERKEMEELDTLNIRPLVKVYKGRHSRTM 722
Db 820 DGMRPEAIRQDPTRKGSVNVNPTNTR-----POSTDPEIRK-----YKRFNSEIL 866

QY 723 IREANFWANFVMSGDCGHIFWDRHTAEHLMLLE--ADNHVNCVLOPHPPDPILASSG 780
Db 867 C--AALWGNLVV-GTESG-----IMLLDRSGGQKVPYPLINRRRFOQMDVLEG 911

QY 781 IDYDIKWSPLEESRIFNRKLADEVITRNELMLEETNTITV 822
Db 912 LNVLTISGKKDLRYVYLSWLRNKLHNDPEVEKKQGTWTV 953
```





```
; FILING DATE: 25-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Treccartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1848 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-839-996-6

Query Match          3.9%; Score 175.5; DB 9; Length 1848;
Best Local Similarity 21.6%; Pred. No. 0.018;
Matches 106; Conservative 68; Mismatches 218; Indels 99; Gaps 19;

QY 250 NKKSCRV-----TSLCYSDGQILVSYSDYIYLFDPKDDTARELKTPTSAEERR 299
Db 920 NKKSNKVVVNSATGFTLQVADTKGEP-----NNHLLTFEDASNATNNLEVTLANGSV 974

QY 300 BELRQPPVKRLRLRGDWS-----DTGPR---ARPESEERDGEQSPNVSLMQMSDMLSRW 352
Db 975 D-----RGWKYKLRNVNGRYDLNPEVEKRNQVTDTTNTITPNDI-----1015

QY 353 FEEAQAQNRGRSRPRGTSQSDISTL--PTVPSSPDLEVSETAWEVDTPAEQFLQ 410
Db 1016 -QADAPSQSN-----NEELARVETVPPPPAPATESAIAEQPETPRAETAQ 1061

QY 411 P-----STSTMSGAQAHSSTSPESPHSTPLLSQSPDSEQRQSVASGHHHTHQSDNNNE 464
Db 1062 PAMEETANTSTETAPKSDTATQTENPSE---SVPSSETTEKVAENPPQENETVAKNQEQ 1118

QY 465 KLSKPKGTGE-----PVLSLHYSTEGTITIKLNFDEWSSASSRSGIGSHCKSGEQ 518
Db 1119 ATEPTQNGEVAKEQDPTVEANTQNEATQSEBK---TEETQTATKSEPTSESVTSNQ 1175

QY 519 EESVPOSSVQPPGDSSTKAPESESSEVDVTKYQGVSAENPVENHINITQSDKFTAK-PL 577
Db 1176 PEKTVSQSTEDKVVVEKEAKVE--TEETQAPQVTSKEPP-----KQAPPAPEVPT 1227

QY 578 DSNGERNDLNRSCGVPESASSEKAKEPETSDQTSATESATNNNTNPPQFQTEATG 637
Db 1228 DTNAEEAQAALQQTQPTTAAAEFTTSPNSKPAETQOPS-----EKTNAEP-----VT 1274

QY 638 PSAHEETSTRDSALQDQDSDDDPVL-IPGARYRAGPCDRBSAVARIOEFFFFRRKERKEM 696
Db 1275 PVVSENTATQTEETAKVEKTEQVQVQVQSPKQEQPA-AKPAQATKQPAEPARE 1333

QY 697 EELDTLNIIRP 707
Db 1334 NVLTKNVGP 1344

RESULT 17
US-09-815-242-5635
; Sequence 5635, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
```

```
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5635
; LENGTH: 2368
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5635

Query Match          3.8%; Score 174; DB 10; Length 2368;
Best Local Similarity 21.6%; Pred. No. 0.03;
Matches 103; Conservative 61; Mismatches 181; Indels 132; Gaps 19;

QY 281 DKPDQDARELKTPTSAEERREELRQPPVKRL--RLRGDWSDTGPRARPESEERERDGSQSPN 338
Db 1961 DNNDATANS-----NANATPENTGQPNVTSTONANATSTSTTTNNQNDAAATGETTATSAN 2016

QY 339 VSLMQMSDMLSRWFEAESEVAQSNRGRGR-SRPRGTSQSDISTLPTVPSSPDLEVS 397
Db 2017 SSAT-----DDANDKPOANNSSADTSTNSPTMDNDVTSKPEVESTNNGTTDKP 2065

QY 398 AMEVD--TPAEQFLOPSTSTMSQAQHSSTSPESPHSTPLLS-----439
Db 2066 ATEADNATPAES-----ATNNSTTTATNENAPTETATATPTASTGAESSADSKDNASVN 2121

QY 440 -----PDSEQRQSVASGHHHTHQSD--DNNNEKLSKPKGTGE 474
Db 2122 DSKQNAEVNNSAESOSTNGKVAQPKSENKAKAEKGRDSTNQSMVESTETL-PSADITE 2180

QY 475 PVLSLHYS--TEGTTTSTIKLNFDEWSSIASRSGIGSHCKSGEQEESFVQ--SSVQP 530
Db 2181 PKVSSNTSKDEESTTSQTDABQHNSDINVASNEAD-----KSEGNVDVDSNKPSTSKP 2235

QY 531 PEGDSETKAPESESE-----DVTYKQGVSAENPVENHINITQSDKFTAKPL----DSN 580
Db 2236 SEAKDKATSTEDSQADMATADTKDNQASIGATADVNN--KATQDGNANASPATVSKGSN 2293

QY 581 SGERNDLNRSCGVPESASSEKAKEPETSDQTSATESATNNNTNPPQFQTEATGPSA 640
Db 2294 SANQDMLNVTK-----TKENKANAKSAQOGKVNKPKQ 2325

QY 641 HEETSTRDSALQDQDSDDDPVLIPGARYRAGPCDRBSAVARIOEFFFFRRKERKEME 697
Db 2326 QAKT-----LPDTGMSHND--LPYAEIALGAG-----MAFLIRRTTKDQO 2365

RESULT 18
US-09-815-242-12389
; Sequence 12389, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
```



```

; APPLICANT: Sim, Kim L.
; TITLE OF INVENTION: Anti-Plasmodium Compositions and Methods of Use
; FILE REFERENCE: 05213-0465 43170-262105
; CURRENT APPLICATION NUMBER: US/09/924,154
; CURRENT FILING DATE: 2001-08-07
; PRIOR APPLICATION NUMBER: US 60/223,525
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 1421
; TYPE: PRT
; ORGANISM: Mammalian
; US-09-924-154-13

Query Match      3.7%; Score 169; DB 10; Length 1421;
Best Local Similarity 19.1%; Pred. No. 0.031;
Matches 136; Conservative 94; Mismatches 241; Indels 240; Gaps 32;

QY 160 CGEDGT-----VRFD-----TRIKTSCKEDCKDDILINCRRAATSA- 198
Db 615 KREDIENIQFRFSEMGDDYQDKTKMIETLKVCEKPCDD---NCKRKCNSYKE 671
QY 199 -ICPPIPYLVGCGSDSSVRIYDRMLGTRATGNAGRTTGMVARE-----IPSHLNKS 253
Db 672 WISKKEEY-----NKQAKQYEQKG-----NNYK-----MYSEFKSIKPEVILKKYS 715
QY 254 CRVTSICYSEGGQELIVSYSDY-----YLFDPKDDTARELKTPSAEE-- 297
Db 716 EKCSNLNFEDEFKEEL---HSDYKNKTCMCPVKDVPISIRNNQTSQEAAPPESTEIA 772
QY 298 REELRQPPVKRLRLRGDWSGTGPARESERERCEGSPNVLQMDSML-----SR 351
Db 773 HRTETR-----DTERKNQEPANKDLKNPQOSVGEN---GTRKDLLEQDLGGR 816
QY 352 WFEASEVAQNRG--RGRSR-----PRGQTSQSDIST-----LPTVP 387
Db 817 SEDEVTOEFGVNHGIPKGEDQTLGKSDAIPNIGPETGISITEESRHEGHKNQALSTSV 876
QY 388 SSPDLEVSETAMEVDTPAEQFLOPSTSTMSAQAHSTSPSPHST--PLLSSPDSEQ 444
Db 877 DEP--ELSDTLQLHEDTKENDKLPLESTI-----TSPTES-----GSSDTEPTS 920
QY 448 VE--ASGHHTHDSNNNEKLSKPKCTGEPVLSLHYSTEGTTSTI----- 491
Db 921 ISEGPKNQEKRRDDLSKLSVSPNSRP-----ETDAKDTNLLKLGVDVISMPEKA 974
QY 492 -----KLNFTDWSIASSSRGIGSHCKSEGOEESFVQSSVOPPEGDSKAPES 543
Db 975 VIGSPNDNINVTQGDNIS---GVNSKPLSD-----DVRPKNHEEVKEHTSN 1020
QY 544 SEDV-----TKYQGVSAENPVNHNITQSKFTAKPLDSN 580
Db 1021 SDNVQSGGIYNNMVEKLDLTENPSSLDGKAHELSNPILSSDDQDMNTPOPLDNT 1080
QY 581 SGE-----RNDLNLDRSCG-----VPEBSASSEKAKEPETSQDTSATNENN 624
Db 1081 SEETTERISNNEYKVNREGERTTKYVEDIVLKSHMNRSDGDELXDENSDLTVDNES 1140
QY 625 TNPEPQFQTEATGSAHEETSTRDSALQDITDSD-----DDPVLIPGARVRAGPGD 675
Db 1141 EDAEAKMKGNDSMSHNSQIHESDQKNDKMTVGDLTGHVQNEISVP----- 1190
QY 676 RRSAVARIQEFRRRKERK-----EMEELDTLNIIRPLVKMYKGHNRSTM 722
Db 1191 ---VTGEIDEXRESKESKIKAHEERLSHTDIIK-----INPEDRNSNTL 1233

RESULT 21
US-09-815-242-10932
; Sequence 10932, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard

```

```

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10932
; LENGTH: 1031
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
; US-09-815-242-10932

Query Match      3.7%; Score 167.5; DB 10; Length 1031;
Best Local Similarity 23.8%; Pred. No. 0.025;
Matches 67; Conservative 48; Mismatches 122; Indels 45; Gaps 9;

QY 387 PSSPDLVEYSETAMEVDTPAEQFLOPSTSTMSAQAHSTSPSPHST--PLLSSPDSEQ 444
Db 718 PAAQKRINQLTQTITALLLVKSTETSTNTSSTSTSTSTSTSTSTSTSTSTSTSTST 777
QY 445 ROSVEASGHHTHHQSDNNNEKLSKPKGTGEPVLSLHYSTEGTTSTIKLFTDEWSSIAS 504
Db 778 STSSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 827
QY 505 SSRGIGSHCKSEGOEESFVQSSVOPPEGDSKAPESSEEDVTYQGVSAENPVNHN 564
Db 828 SS-----TSSTTNSSSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 879
QY 565 NITQSDKFTAKPLDSNGERNDLNLDRSCGVPEESASSEKAKEPETSQDTSATNENN 624
Db 880 SET-----SNTNESNTPSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 928
QY 625 -----TNPEPQFQTEATGSAHEETSTRDSALQD 653
Db 929 QSKGNSVIYAVESNQDPNDAQSNKPSA-KASQTKESVAEN 969

RESULT 22
US-09-815-242-13057
; Sequence 13057, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard

```





Db 676 KKEYNKQAKQYQKGNMYSEKSIKPEVYLKKYSEKSNLNFDEFKEE-----730  
QY 374 GTSQSDISTLPVPSPLEVESETAMEVDTPAQFLQPTSTSTMSAQAHSTSP--TESP 431  
Db 731 --LHSDYKKNKIM--CP--EVKDVPIIIRNNEQTQSEAVPEENTIAHRTSPISSEGP 784  
QY 432 HSTPLSSPDSEQRQVEASGHHTHQSDNNNEKLSKPKFGTGPVLSLHYSTEGTTTSTI 491  
Db 785 -----KGNEKQERDDDSLKISVSPNSRP-----ETDAKDTNL 819  
QY 492 -----KLNFTDEWSSSTASSRSGIGSHCKSGEGEESFVPOSSVQ 529  
Db 820 LKLGVDISMPKAVIGSPNDININTEQDNLS-----GVNSKPLSD-----DVR 865  
QY 530 PPEGSETRKAPESSESDVTKYQEGVSAENPVENHINITQSDAKFTAKPLDSNSG-----ER 584  
Db 866 PDKKELEDQNSDESETVYH--ISKSPSINNGDDSGSATVSSSSSNTGLSIDDDR 922  
QY 585 NDLNDRSCGVPEESASSEKAKEPETSDOTSTES-----ATNENNTNPEQFOFTEATG 637  
Db 923 NGDTFVRT-----QDANTEDVIRKENADKDEKGADEERHSTSESLSPEEKMLTDNEG 978  
QY 638 PSA--HEE-----TSTRDSA-----LQDT-----DD-----656  
Db 979 GNSLNEEVEKHTSNDVQSGGIVNMVKEKLDLENPSSLDDEGAHEELSEPNLS 1038  
QY 657 SDDPVLIPGARYACGDRRSARVARIQEFFRR--RKERKEMEELDTLIRPLVKMYK 714  
Db 1039 SQDMSNTP-----GPLDNTS-----EETTERISNNEKYKVEREDERTLTKEYEDIVLK 1087  
QY 715 GHNSR-----THIKANFWGANFVMSGDCGHIFWDRITABHLMLEAD 760  
Db 1088 SHMRESDDGELYDRNSDLSTVNDSE--DAEAKMKGNDTSEM--SHNSQH-----IESD 1139  
QY 761 -----NHVYNCLQPHFPDPTLASSGDIYDI-----KWSPLEESRIFN--798  
Db 1140 QKNDKMTVGLDGTTHVQNEIS-----VPVTGEIDELKRESKSI--HKAEEERLSHTD 1192  
QY 799 --RKLADDEVITRNLMLLETRN 818  
Db 1193 IHKINPDRNSNTLHKDIRN 1213

## RESULT 24

US-10-025-380-1065  
; Sequence 1065, Application US/10025380  
; Publication No. US20020182191A1  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Secrist, Heather  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Meagher, Madeleine Joy  
; APPLICANT: Stolk, John A.  
; APPLICANT: Wang, Tongtong  
; APPLICANT: Jiang, Yugu  
; APPLICANT: Smith, Carole L.  
; APPLICANT: King, Gordon E.  
; APPLICANT: Wang, Aijun  
; APPLICANT: Clapper, Jonathan D.  
; APPLICANT: Skeiky, Yasir A. W.  
; APPLICANT: Fanger, Gary R.  
; APPLICANT: Vedick Thomas S.  
; APPLICANT: Carter, Darick  
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS  
; TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE  
; FILE REFERENCE: 210121.471C14  
; CURRENT APPLICATION NUMBER: US/10/025,380  
; CURRENT FILING DATE: 2001-12-19  
; NUMBER OF SEQ ID NOS: 1129  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1065

; LENGTH: 957  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-025-380-1065  
Query Match 3.6%; Score 164.5; DB 9; Length 957;  
Best Local Similarity 19.9%; Pred. No. 0.035;  
Matches 132; Conservative 81; Mismatches 301; Indels 149; Gaps 28;

QY 126 EQDAENR--QCQFCHYGYTYEIMVFN-----DPYFLSC--GEDGVVWFDTRIKT- 175  
Db 204 ESDTTSGRGEESTSHSTHTTSSAPSTTSALVEEPTSYHSPGSTATTHTFPDSSTSG 263  
QY 176 -----SCTKEKDDKILLINCRRAATVAICPPVYLAAGCSDSVRIYDRMLGTTRA 228  
Db 264 RSESTASHSNQDATGIVLPA--RSTSVLLGEGTSPISSSGSMETTA-----LPGSTT 316  
QY 229 TGNVAGRGTTGMVARIPSHLNKSCRVT-----SLCYSEDGQELIYSYS---SDIYILFD 281  
Db 317 TPLGSEKSTT-----FHSSPRSPATTLSPASTTSSGVSEESTTSHSRPGSTHTTAF- 367  
QY 282 PKDDTARELKTPSAERREELRQPP--VKRLRLRGWSDTGPRAPESEERERGEQSPNV 339  
Db 368 --PDSIT--TPGLSRHSTSHSPGSTDTTLLPASTTTSGP--SQESTTSHSPGSTDT 420  
QY 340 SLMQMSDMLSRWFEEASEVAQSNRG-----RGRSRPRGGTQSQDIST 382  
Db 421 ALSPGSTTALS--FGOESTTFHSSPGSTHTTLFPDSTTSSGIVEASTRVHSSTGSPRTLL 478  
QY 383 LPTVPSPPLEVSETAMEVDTPAEQFLQPTSTSTMSAQAHSTSPTESPSTPL-----437  
Db 479 SPASSTSPGLQESTAFQTH--PASTHTTPTSTATAPVEESTTYHRSPTSSTTTHFPAS 537  
QY 438 -----SSPD-----SEQRQVEASGHHTHQSDNNNEKLSKPKG--TCEPV 476  
Db 538 STTSGHSEKSTTFHSPSPDASGTTTPSSAHSTTSGRGESTTSRISPGSTEITTLFGSTTTPG 597  
QY 477 LS-----LHYSTEGTTTSTIKLNFDEWSSSTASSRSGIGSHCKSE---GOEEFVPOSSV 528  
Db 598 LSEASTTFYSSPRSPPTTL-----SPASMTSLGVEESTTSRQPGSTHTVSPAST 649  
QY 529 QPP-----EGDSETKA--PEESEDVTKYQEGVSAENPVENHINI-----566  
Db 650 TTPGLSEESTVYSSSPGSTETTVFPRSTTSYVRGEPTTFHSPASTHTLTFTEDSTTS 709  
QY 567 ---TQSDKFTAKPLDSNGERNDLNL-----DRSCGVPEESASSEKAKEPETSDOTSTESA 619  
Db 710 GLTEESTAPPGSPASTQTGLPATLTADLGEESTTTPSSSGSGTGTTLSPARSTTSLVGE 769  
QY 620 TNENNTNPEQFOFTEATG-----PSAHEETSTRDSALQDSTDSDDPVLIPGARYRAGPD 675  
Db 770 STPSRLSPSSTETTLPGSPPTTSLSEKSTT-----FYTSRSPDATLSPATTTSSGVSE 824  
QY 676 RRS 678  
Db 825 ESS 827

## RESULT 25

US-09-922-217-1065  
; Sequence 1065, Application US/09922217  
; Patent No. US20020076414A1  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Secrist, Heather  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Meagher, Madeleine Joy  
; APPLICANT: Stolk, John A.  
; APPLICANT: Wang, Tongtong  
; APPLICANT: Jiang, Yugu  
; APPLICANT: Smith, Carole Lynn  
; APPLICANT: King, Gordon E.



Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 5816  
LENGTH: 2478  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-09-815-242-5816

Query Match 3.6%; Score 164.5; DB 10; Length 2478;  
Best Local Similarity 21.1%; Pred. No. 0.12;  
Matches 97; Conservative 71; Mismatches 187; Indels 105; Gaps 21;  
QY 281 DPKDDTARELKTPSAERREELRQPPVKRLRLGDSMTGPRAPRESERERDGEQSPNVS 340  
DB 2078 DNNDAFANS-----NANATPENTGQPNVSETTANGK-ADASP-TTPNNSDAATGETTATSA 2131  
QY 341 LMQRMSDMLSRWFEEASEVAQSNRGRG-RSRPRGGTSQSDISTLPTVPSPDLEVSETAM 399  
DB 2132 T-----DDANDKPQANNSSVDASTNSPTMDNDVTSKPEVESTNNGTTDKPVT 2179  
QY 400 EVD--TPAQFLOPSTSTMSAQAHSTSPESP--HSTPLLSPDSEQSQVEASGHHT 455  
DB 2180 ETDNATPAESTTNNSTTTATNENAPTGTATAPTASTAEASADSKDNASVNDKQNA 2239  
QY 456 H---HSDNNNEKLSKPKGTPGVLSLHYSTGTTTSTIKLNFTEWSSIASSSRGIGS 511  
DB 2240 EYVNSAESOSTNDKVA-QPKSENKAKAEKDGSDSTNQSWVE-STTETLPADIETPNVPS 2297  
QY 512 HCKSEGESEFVPOSSV-----QPPEGDSE-----TKAPEESSEDVTK 549  
DB 2298 N-TSKDKKEESTTQTDAGOLKSETNVASNEADKSPSKADTEVSNKPTSSASSEAKEKMTS 2356  
QY 550 YQEGVSAENPVENHINITOSDKFTAKPLDSNGERNDLNRSCGVPPEESASSEAKEPE 609  
DB 2357 -----TNVSKD-----DTATADTND--TQKSVG-----SAANKATQ-- 2387  
QY 610 TSDQSTESATNNTNPEQFQTEATGSAHE-----ETSTRDSALQDITDSDS 657  
DB 2388 -NDGANASPATVSNNGNSANQDMLNVTNTDDHOAKTKSAQQGKYNKAKQOAKTLPTDGM 2446  
QY 658 DDDPVLPGARYAGCDRRSAVARIQEFFFFRRKERKEME 697  
DB 2447 HND--LPYAEALGAG-----MAFLIRRTTKDQ 2475

RESULT 28  
US-09-815-242-12967  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 12967  
LENGTH: 2478  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-09-815-242-12967

Query Match 3.6%; Score 164.5; DB 10; Length 2478;  
Best Local Similarity 21.1%; Pred. No. 0.12;  
Matches 97; Conservative 71; Mismatches 187; Indels 105; Gaps 21;  
QY 281 DPKDDTARELKTPSAERREELRQPPVKRLRLGDSMTGPRAPRESERERDGEQSPNVS 340  
DB 2078 DNNDAFANS-----NANATPENTGQPNVSETTANGK-ADASP-TTPNNSDAATGETTATSA 2131  
QY 341 LMQRMSDMLSRWFEEASEVAQSNRGRG-RSRPRGGTSQSDISTLPTVPSPDLEVSETAM 399  
DB 2132 T-----DDANDKPQANNSSVDASTNSPTMDNDVTSKPEVESTNNGTTDKPVT 2179  
QY 400 EVD--TPAQFLOPSTSTMSAQAHSTSPESP--HSTPLLSPDSEQSQVEASGHHT 455  
DB 2180 ETDNATPAESTTNNSTTTATNENAPTGTATAPTASTAEASADSKDNASVNDKQNA 2239  
QY 456 H---HSDNNNEKLSKPKGTPGVLSLHYSTGTTTSTIKLNFTEWSSIASSSRGIGS 511  
DB 2240 EYVNSAESOSTNDKVA-QPKSENKAKAEKDGSDSTNQSWVE-STTETLPADIETPNVPS 2297  
QY 512 HCKSEGESEFVPOSSV-----QPPEGDSE-----TKAPEESSEDVTK 549  
DB 2298 N-TSKDKKEESTTQTDAGOLKSETNVASNEADKSPSKADTEVSNKPTSSASSEAKEKMTS 2356  
QY 550 YQEGVSAENPVENHINITOSDKFTAKPLDSNGERNDLNRSCGVPPEESASSEAKEPE 609  
DB 2357 -----TNVSKD-----DTATADTND--TQKSVG-----SAANKATQ-- 2387  
QY 610 TSDQSTESATNNTNPEQFQTEATGSAHE-----ETSTRDSALQDITDSDS 657  
DB 2388 -NDGANASPATVSNNGNSANQDMLNVTNTDDHOAKTKSAQQGKYNKAKQOAKTLPTDGM 2446

QY 658 DDDPVLIPGARYRAGGDRRSASAVARIQEFRRKRRKEME 697  
Db 2447 HND--LPYAEALGAG-----MAFLIRRTKKDQ 2475

RESULT 29

US-09-765-272-94  
; Sequence 94, Application US/09765272  
; Patent No. US20020061545A1  
; GENERAL INFORMATION:  
; APPLICANT: Choi et. al.  
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
; NUMBER OF SEQUENCES: 452  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/765,272  
; FILING DATE: 22-Jan-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/961,083  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brookes, A. Anders  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: P8340P2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 94:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 278 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 94:

Query Match 3.6%; Score 164; DB 10; Length 278;  
Best Local Similarity 22.2%; Pred. No. 0.0074;  
Matches 66; Conservative 50; Mismatches 117; Indels 64; Gaps 10;  
QY 387 PSSPDLEYSETA-----MEVDTPAEQFLOPSTSTMSAQAHSTSSPTE 429  
Db 9 PESDPTVSDRGEPOQVAPLPEYKGNIEQVRPEPVEKTEKQEPKTEEV-----PVK 61  
QY 430 SPHSTPLSSPDSEQRQVVEASGHITH--HOSDNNKLSKPKGTGEPVLSLHSTEGTT 487  
Db 62 PTEETPVNPNEGTTTGTSTGIEQAEVNPVPAEESTTNSKRVSP-----DTSSKN 108  
QY 488 TSTIKLNTDSESSSTASSRGISGCKSEGOEESFPVPOSSVQPPGDSSETKAPESSESDY 547  
Db 109 TGEVSSNPSSDTSVGSNGKPEHNDKSNSEKT--VEEVPVNPNEGTVGEGTSNQETKPV 167  
QY 548 TKYQGVSAENPVENHINITQDKFTAKPLDSNSGNGKDLNDRSCGVPESASSEK--- 604  
Db 168 QPAEE-----TQTN--SGRIANETGEVSNKPSDSKPPV-EESNQEPKNGT 210  
QY 605 AKPEPSTQSTESATNNNNPNPEQFOTATGSPSAHEETSTRSALQDSTDSDDDP 661  
Db 211 ATKPENSNGTTSN-----GQTEPEPSNGNSTEDVSTESNTSNGNEEIKQENELDP 263

RESULT 30

US-10-077-111-10  
; Sequence 10, Application US/10077111  
; Publication No. US20020187492A1  
; GENERAL INFORMATION:  
; APPLICANT: Todderud, C. Gordon  
; APPLICANT: Finger, Joshua N.  
; APPLICANT: Rillema, Jill  
; TITLE OF INVENTION: TBA  
; FILE REFERENCE: 3053-4114US2  
; CURRENT APPLICATION NUMBER: US/10/077,111  
; CURRENT FILING DATE: 2002-02-15  
; PRIOR APPLICATION NUMBER: 60/294,181  
; PRIOR FILING DATE: 2001-05-29  
; PRIOR APPLICATION NUMBER: 60/269,366  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 10  
; LENGTH: 1356  
; TYPE: PRT  
; ORGANISM: Podospora anserina  
; FEATURE:  
; OTHER INFORMATION: beta transducin-like protein encoded by the  
; OTHER INFORMATION: het-e-1 gene  
US-10-077-111-10

Query Match 3.5%; Score 160.5; DB 9; Length 1356;  
Best Local Similarity 20.9%; Pred. No. 0.096;  
Matches 97; Conservative 65; Mismatches 182; Indels 121; Gaps 21;  
QY 46 TLNVHDCVNTICWNTDTEYILSGSDDTKLVISNPSYRKVLFTIRSHRANIFSAKFLPC 105  
Db 920 TLEHGGRVQSVAFSPDQGVASGSDHTIKIWAASGCTQTL-EGHGSSVLSVAFSP- 977  
QY 106 TNDKQIVSCSGDGVIFYTNVEQDAETNQCQFTC-----HYGTYIEMTVNDPYTFLS 159  
Db 978 -DQQRVASGSGDKTIKIWDTAGS-----TCTOTLEHGGSVMSVAFSP-DGQRVAS 1026  
QY 160 CEDGTVRFDFTRIKTSCTKEDCKDDILINCRRAATVAICPPPIPYLAVGSDSSVRIY 219  
Db 1027 GSDDKTIKIWDTAGSCT-CTQ-----TLEHGGSVQSVFSPDQGVASGSDHTIKIW 1078  
QY 220 DRMLGTRATGNAGRTTGMVAREIPSHLNKSCRVTSLCYSESDGQELVLSYSDIYL 279  
Db 1079 D-AVSGT-CTOTLEHGGS-----VMSVAFSPDQGVASGSDGTIKI 1119  
QY 280 FDPKDDTARELKTPSAERREELRQPPVKRLRLGDSDTGPRARPESEREDGEQSPNV 339  
Db 1120 WDAASGTC-----TQTLGEGGWVHS-VAFSPDQGVASGSDGTI 1159  
QY 340 SLMQMSDLSRWFEEASEVAQSNRGRSRPRGTSQSDISTLTPTVPSSPDEVSETAM 399  
Db 1160 KI-----WDAASGTCQTLGEG-----HGWSV-----VAFSPD----- 1188  
QY 400 EVDTPAEQFLOPSTSTMSA--QAHSTSPSTPSTPLLS---SPDSEQRQSVASGHH 454  
Db 1189 -----QORVASGSSDKTIKIWDTAGSCTQTLGEGGWVQSVAFSPDQGV-----VASG-- 1237  
QY 455 THQSDNNKLSKPKGTGEPVLSLHSTEGTTTSTIKLNTFDEW 499  
Db 1238 ---SSDNTIKIWDTAGSCTQTLNV-----GSTATCLSFDTNAY 1274

RESULT 31

US-09-815-242-5904  
; Sequence 5904, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlson, Kari L.  
; APPLICANT: Zyskind, Judith W.

APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
TITLE OF INVENTION: Prokaryotes  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIORITY FILING DATE: 2000-03-21  
PRIORITY FILING DATE: 2000-03-21  
PRIORITY FILING DATE: 2000-05-23  
PRIORITY FILING DATE: 2000-07-27  
PRIORITY FILING DATE: 2000-05-26  
PRIORITY FILING DATE: 2000-04-28  
PRIORITY FILING DATE: 2000-10-23  
PRIORITY FILING DATE: 2000-11-27  
PRIORITY FILING DATE: 2000-12-22  
PRIORITY FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 5904  
LENGTH: 502  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-09-815-242-5904

Query Match 3.5%; Score 160; DB 10; Length 502;  
Best Local Similarity 21.9%; Pred. No. 0.038;  
Matches 85; Conservative 56; Mismatches 148; Indels 100; Gaps 12;

QY 312 LRGWSDTGPAPRPERERDEQSPNVLQRMDSMLRWFEAESEVAQSNRGRSRP 371  
Db 53 LSGTSESESDTSSESKSD-STSMISMSQSTS----- 86  
QY 372 RGGTQSQSDISTL-PTVPSSPDLESETAMEVDT-PAEQFLQPTSTSTMS----- 418  
Db 87 -GSTSTSTSLSDSTSTSLASMSQGVDSNASASASSTSTSTSESQSTSSYT 145  
QY 419 ----AQHSTSPSPSTPLSGSPDSEQRQSVASGHHTHQSDNNNEKLSRPGTGE 474  
Db 146 SQSTQSSESTSTSLSDSTSTSLKSTSGSVSTGASLGSSESDSISSTSESTSE 205  
QY 475 PVL-----SLHYSTEGTTSTIKLNTDEWSSIASSSRGIGSHCKSEGESEFVQSSVQ 529  
Db 206 SASTSLSDSTSTSGSASTSTSLN-----NSASASEDLSSTSLSDS-----TSASMQ 254  
QY 530 PPEGDSSTKAPES-----SEDTKYQEGVSAENPVENHINITQSDKF 572  
Db 255 SSESQSTASLSDSTSTSNRMSTTASLSTSVSTSEGSTSE-----STSESST 307  
QY 573 TAKPLDNGRNDLNRDCGVPPEASSEKAKBPETSDQSTSESATN-----ENNTNPE 628  
Db 308 STSLSDSQSTSRK-----TSASGASTSTSTSDSRSTASSTSMRTSTSDSQ 355  
QY 629 PQFQTEATGPAHETSTRDALQDSDS 657  
Db 356 SMSLSTSTSTSMDSSTSLSDSVSDSTSDS 384

RESULT 32

US-09-801-368-408  
Sequence 408, Application US/09801368  
Patent No. US20020128250A1  
GENERAL INFORMATION:  
APPLICANT: Busby, Robert  
APPLICANT: Cali, Brian  
APPLICANT: Hecht, Peter

APPLICANT: Holtzman, Doug  
APPLICANT: Madden, Kevin  
APPLICANT: Maxon, Mary  
APPLICANT: Milne, Todd  
APPLICANT: No. US20020128250A1man, Thea  
APPLICANT: Royer, John  
APPLICANT: Salama, Sofie  
APPLICANT: Sherman, Amir  
APPLICANT: Silva, Jeff  
APPLICANT: Summers, Eric  
TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi  
FILE REFERENCE: 109272.147  
CURRENT APPLICATION NUMBER: US/09/801,368  
CURRENT FILING DATE: 2001-03-07  
PRIORITY FILING DATE: 2000-03-07  
PRIORITY FILING DATE: 2000-01-19  
PRIORITY FILING DATE: 1999-10-20  
NUMBER OF SEQ ID NOS: 440  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 408  
LENGTH: 713  
TYPE: PRT  
ORGANISM: Saccharomyces cerevisiae  
US-09-801-368-408

Query Match 3.5%; Score 156.5; DB 10; Length 713;  
Best Local Similarity 19.6%; Pred. No. 0.072;  
Matches 111; Conservative 92; Mismatches 225; Indels 137; Gaps 23;

QY 286 TARELKTSPAEERELRQPPVKRLRLRGDWSDTG-PRARPESERERDEQSPNVLQMR 344  
Db 65 THRKMKDAYEEE-----IKHLKLEQRDHOIASLTVOOQROOQOQOQVQHLQOQ 115  
QY 345 MSDMLSWFEAESEVAQ-----SNRERG-----RSRPGGTQSQSDISTL 383  
Db 116 QQQLAAA--SASVVAQOQPPATTSATATPAATTTGSPSAFPVQASRPRLVGSQPTTL 173  
QY 384 PTVPSSPDLESETAMEVDTPAEQFLQPTSTSTMSAQAHSTSSPTSEPHST-PLLSPPDS 442  
Db 174 PVVSSNAQQLPQOQLOOQOQPPQVSVAPLSNTAINGSPSTKETTTLPSVKAPES 233  
QY 443 EQQSVASGHHTHQSDNNNEKLSRPGTGEVLSLHYSTEGTTST-IKLNTDEWSS 501  
Db 234 TLKET-----EPENNTSKIN--DTGSA-----TTATTTTATETETKPEEDAT 275  
QY 502 IASSRGIGSHCKSEGESEFV-----QSSVQPP-----EGDSEI--KAPESESDYTK 549  
Db 276 PASLH-----QDHYLVYNQRANHSKPIPPFLDLDSQSVDPALKQTNDYIL 324  
QY 550 YQEGVSAENPVE-----NHINITQSDKFTAKPLDNGRNDLNRDCGVPPEASSEK 604  
Db 325 YNPALPREIDVELHKLSDHVSVCVKF-----SNDEYLATGCNKTKTYRVSDGSLV 378  
QY 605 AKETPESDQSTSEATNNTNPEQFQTEATGPAHETSTRDALQ-----DTDD 656  
Db 379 ARLSDDSAANHRNSITENTTTSTDNNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNT 438  
QY 657 SDDDPVLI-----PGARYA-GPDRRSARVATQEFFRKRKEMBEELDTLNRRLPV 709  
Db 439 SPSSDLYIRSVCFSPDGKFLATGAEDR---LIRLWDIENRK-----I 477  
QY 710 KMVYKGRNSRTMIKEANFWGANF-----VMSGSDCHGIFWDRHTAEHLMLLEADNHV 763  
Db 478 VMILQGH-----EQDIYSLDYFPPSGDKLVSGSGDRTVRWLDTLQCCSLTSLIEDGV 529  
QY 764 VNCLOPHFPDPIASSGIDYDIKIW 788  
Db 530 TTVAVSPDGKYGIAAGSLDRAVRVM 554

RESULT 33

US-09-820-843a-12

; Sequence 12, Application US/09820843A  
; Publication No. US20030039963A1  
; GENERAL INFORMATION:  
; APPLICANT: Council of Scientific and Industrial Research  
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEIN  
; FILE REFERENCE: Q63915  
; CURRENT APPLICATION NUMBER: US/09/820,843A  
; CURRENT FILING DATE: 2001-03-30  
; NUMBER OF SEQ ID NOS: 118  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 12  
; LENGTH: 505  
; TYPE: PRT  
; ORGANISM: H. pylori  
; NAME/KEY: misc feature  
; OTHER INFORMATION: poly E-rich protein  
; NAME/KEY: misc feature  
; OTHER INFORMATION: g1/2313421  
US-09-820-843A-12

Query Match 3.4%; Score 155.5; DB 9; Length 505;  
Best Local Similarity 18.2%; Pred. No. 0.053;  
Matches 74; Conservative 73; Mismatches 152; Indels 107; Gaps 15;

QY 239 GMVARFIPSHLNKSCRYTSLCYSDGQILVSVSSDYIYLFDPKDDTARELKTSPSABER 298  
DB 145 GDLEALVQEPNNEBQLPTLNDQEEKEVKE-----EKEEVEKEEKEVKEE 194  
QY 299 REELRQPPVKRLRGDWSDTGPRARP----- 325  
DB 195 KEVKETQEEKKPKDDTQGETLKDKEVSKELPAQELPKEETQPDPIKEETOEN 254  
QY 326 ESERERDGEQSPNVSLMQMSDMLSRWFEEASEVAQSNRGRSRPRGCTSQSDISTLPT 385  
DB 255 KEKQEKTDSPSAQELAMQELV-----KEIQENSGQ-----ENKKTQES 297  
QY 386 VPSPDLVSET-----AMEVDTPAEQFLOPSTSTMSAQAHSTSP--TESPHSTPL 437  
DB 298 AEIPQDKIEQVVTETKTAQAELEVPKEK--TQESAEALQETAHELEKQAEIAETPDQVEIP 356  
QY 438 SSPDSEQSQVSEASGHHTHQSDNNNEKLSKPGTGEPVLSHYSTEGTTSTIKLNTD 497  
DB 357 QSQDKEV-QELEIPKEETOENTETPDVETPQE-----KETQED 394  
QY 498 EWSSTASSRIGSGHCKSEGESEFVPOSSVQPPGDSSETKAPESSESDVTKYQEGVSAE 557  
DB 395 HYESIEDIPEPV--MAKANGEELPFLNEAVAKIPNNENDTETPKESVTETSKNEN--NTE 450  
QY 558 NPVENHINITQSDKFTAKPLDSNGERNDLNL--DRSCGVPEESASS 602  
DB 451 TPQEK-----ESDK--TSSPLELR-----LNLQDLLKSLNQESLKS 485

## RESULT 34

US-09-815-242-11317  
; Sequence 11317, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyglind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11317  
; LENGTH: 505  
; TYPE: PRT  
; ORGANISM: Helicobacter pylori  
US-09-815-242-11317

Query Match 3.4%; Score 155.5; DB 10; Length 505;  
Best Local Similarity 18.2%; Pred. No. 0.053;  
Matches 74; Conservative 73; Mismatches 152; Indels 107; Gaps 15;

QY 239 GMVARFIPSHLNKSCRYTSLCYSDGQILVSVSSDYIYLFDPKDDTARELKTSPSABER 298  
DB 145 GDLEALVQEPNNEBQLPTLNDQEEKEVKE-----EKEEVEKEEKEVKEE 194  
QY 299 REELRQPPVKRLRGDWSDTGPRARP----- 325  
DB 195 KEVKETQEEKKPKDDTQGETLKDKEVSKELPAQELPKEETQPDPIKEETOEN 254  
QY 326 ESERERDGEQSPNVSLMQMSDMLSRWFEEASEVAQSNRGRSRPRGCTSQSDISTLPT 385  
DB 255 KEKQEKTDSPSAQELAMQELV-----KEIQENSGQ-----ENKKTQES 297  
QY 386 VPSPDLVSET-----AMEVDTPAEQFLOPSTSTMSAQAHSTSP--TESPHSTPL 437  
DB 298 AEIPQDKIEQVVTETKTAQAELEVPKEK--TQESAEALQETAHELEKQAEIAETPDQVEIP 356  
QY 438 SSPDSEQSQVSEASGHHTHQSDNNNEKLSKPGTGEPVLSHYSTEGTTSTIKLNTD 497  
DB 357 QSQDKEV-QELEIPKEETOENTETPDVETPQE-----KETQED 394  
QY 498 EWSSTASSRIGSGHCKSEGESEFVPOSSVQPPGDSSETKAPESSESDVTKYQEGVSAE 557  
DB 395 HYESIEDIPEPV--MAKANGEELPFLNEAVAKIPNNENDTETPKESVTETSKNEN--NTE 450  
QY 558 NPVENHINITQSDKFTAKPLDSNGERNDLNL--DRSCGVPEESASS 602  
DB 451 TPQEK-----ESDK--TSSPLELR-----LNLQDLLKSLNQESLKS 485

## RESULT 35

US-09-864-761-36182  
; Sequence 36182, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO  
; FILE REFERENCE: Aescmca-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26



Db 481 RRG-----SISVSPKP-----LSSFKMSLRNWNVTRTPSSPPV 514  
QY 510 ---GSHCKSEGESESVPOSSVOPPEDGETKAPESSESDVTYKQGVSAENPVENHIN- 565  
Db 515 TTPASETKISSPRKALIPVS-----OKSQADACSESRRNRRLDSCLIESVKQKCVKS 569  
QY 566 ---ITQSDKFTAKPLDSNGERNDLNRSCGVPESASSEKAKEPETSDOISTESATNE 622  
Db 570 CNCVTELD-----GQAESRLDLCCLSGTOGSEVLSDQSEGPTKSKTEGAGTSIS 618  
QY 623 NNTNPEPQTEATGP-----SAHEETSTRD-----SALQDITDSD--D 659  
Db 619 EPPSPVPYASECGPLPLRPGGSEWVGKENSPPENKNWLLAIAAKRAENSPPRS 678  
QY 660 DVLIPGARYRAG---PGD---RRSAVARIQEFRRRKE 692  
Db 679 PSSQTPSSRRSQSKTSPGVTITPSSMRKICTYFRKRKTQ 717

## RESULT 37

US-10-108-605-195  
; Sequence 195, Application US/10108605  
; Patent No. US20020160934A1  
; GENERAL INFORMATION:  
; APPLICANT: Broadus, Julie  
; APPLICANT: Stam, Lynn  
; APPLICANT: Bachmann, Jane  
; APPLICANT: Kamdar, Klm  
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCODE  
; TITLE OF INVENTION: PROTEINS ESSENTIAL FOR LARVAL VIABILITY AND USES THEREOF  
; FILE REFERENCE: 31133B  
; CURRENT APPLICATION NUMBER: US/10/108,605  
; CURRENT FILING DATE: 2002-03-27  
; PRIOR APPLICATION NUMBER: US 09/761,142  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/176,418  
; PRIOR FILING DATE: 2000-01-14  
; NUMBER OF SEQ ID NOS: 361  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 195  
; LENGTH: 980  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
US-10-108-605-195

Query Match 3.4%; Score 152.5; DB 9; Length 980;  
Best Local Similarity 20.5%; Pred. No. 0.19;  
Matches 92; Conservative 70; Mismatches 193; Indels 93; Gaps 17;

QY 256 VTSLC--YEGDGOEIL--VSYSSDXYILDPKDDTARELKTPTSAERREELRQPPVKRLR 311  
Db 599 IITLCVYFVENFSKIFDEIYQTFYKLNRYDQYQDLKDRDKMENRTDGGLPITR--- 555  
QY 312 LRGDWSDTGPRAPESERERDGEQSPNVSLMQMSDMLSRWFPEASEVAQ-----SNRGR 366  
Db 656 -----SGGRF-----RDQRQ-----MEEEREM---WFNEEDDFTEEDITYNNVKM 693  
QY 367 GRSRPGGTQSQSDISTLPTVPSPDLEVSETAMEVDTPAEQFLQPTSTMTMSAQAHSTSS 426  
Db 694 SVSEKNGPQTQNOQKSSPPHSTP-----HSGILGSLSTASTATSTATS 738  
QY 427 --PTESPHTPLLSPPDSBORQSV--ASGHHHTHQSDDNNNEKLSRKPCTGEP--VLSLHY- 481  
Db 739 GAPVAGSSSPPAISADEQTQAAVHLAAALQHQQOQQOQQOQNPPQQTQPEIAELOQO 798  
QY 482 --STEGTTTSTIKLNF-----TDWESSIASSSRIGSHCKSEGESESVPOSSVQPP 531  
Db 799 LSSVEAPOQSOELESQAASASPTSSSSSLRSTSSSSASSSSSSSSSSSSPPGSSAAS 858  
QY 532 EGDSETKAPEESEDVTYKQGVSAENPVENHINTQSDKFTAKPLDSNGERNDLNR 591  
Db 859 LCDSATVAAVAASQFLSTIATAMAA-----SVTAAATNSSP-----SISP 899

QY 592 SCGVPESSASSEKAKEPETSDQITSTESATNNTNPEPQTEATGSPSAHEETSTRDSAL 651  
Db 900 APAVSPDIENADAQLPPSDDASSPASGEQDANS-----TEGTSEADKTTAKRGLVD 952  
QY 652 QDTDDDDDDPVLIPGARYRAGPGDRSA 679  
Db 953 YESDSGEDD---YEEDEYSEGPOAQKRA 977

## RESULT 38

US-09-738-877-3  
; Sequence 3, Application US/09738877  
; Patent No. US20020015970A1  
; GENERAL INFORMATION:  
; APPLICANT: Murray, Richard  
; APPLICANT: Watson, Susan  
; APPLICANT: Weiss, Stephen J.  
; APPLICANT: Glynnne, Richard  
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSIS OF ANGIOGENESIS, COMPOSITIONS, AND  
; TITLE OF INVENTION: SCREENING FOR ANGIOGENESIS MODULATORS  
; FILE REFERENCE: A-69806/DJB/JJD  
; CURRENT APPLICATION NUMBER: US/09/738,877  
; CURRENT FILING DATE: 2000-12-15  
; PRIOR APPLICATION NUMBER: US 09/637,977  
; PRIOR FILING DATE: 2000-08-11  
; PRIOR APPLICATION NUMBER: US 60/148,425  
; PRIOR FILING DATE: 1999-08-11  
; PRIOR APPLICATION NUMBER: PCT/US 00/22061  
; PRIOR FILING DATE: 2000-08-11  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 3  
; LENGTH: 1781  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-738-877-3

Query Match 3.4%; Score 152.5; DB 10; Length 1781;  
Best Local Similarity 20.7%; Pred. No. 0.42;  
Matches 141; Conservative 86; Mismatches 249; Indels 205; Gaps 33;

QY 281 DPKDDTARELKTPTSAERREELRQPPVKRLRGLDWSGTGPRAPESERERDGEQS 336  
Db 339 DGKAEVASEKLTASEQAHPQPAESAHEP-----RLSAEY-----KVLPSEQVSGSQG 389  
QY 337 PNVSLMQMSDMLSRWFPEASEVAQSNRGRSRPRGTSQSDISTLPTVPSPDLEVSE 396  
Db 390 PS---BEKPAPLATEVFEDEKIEVHQEE-----VVAEVHVSVTEERTEQKTEVEE 436  
QY 397 TAMEVDTPAEQFL---QPSTSS-----TMSAQAHSTSPSPHSTP---LLSPDSE 443  
Db 437 TAGSV--PAEELVGMDAEPQAEPAKELYKLTCTCVSGEDPTQGADLSPDEKYLKPPGEG 494  
QY 444 QROSVASGHHHTHQSDDNNNEKLSRKPCTGEPVLSLHYSTEGTTTSTIKLNFDEWSSIA 503  
Db 495 VYSEVEML-----SSQERMKVQ---GSPLKLLF-----TSTGLKLLSGKKQ----- 532  
QY 504 SSSRGTGSHCKSEGESESVPOSSVQPPPE---GDSFTKAPESSE----- 545  
Db 533 KGRKGGD---EESGHTQVPADSPDSBEQKGESSASSPPEPEITCLEKGLAEVQDQG 589  
QY 546 -----DVTYKQGVSAENPVENHINTQSDKFTAKPLDSNGERNDLNRSCGVPE 597  
Db 590 EAEGATSDGEKKREGVT---PWASTFKMVTPKKRVRRP---SESDEDELDKVKYSATLSS 644  
QY 598 -ESASE-----KAKEPETSDOTST-----ESATNNTNPEPQFQ 632  
Db 645 TESTASEQEMKSGVSEEPKPEPKKVTSTVSWEALICVSGSKKARRRSSD----- 698  
QY 633 TEATGESA-----HEETSTRDSALQDITDSD-----DPVLIPGARYRAGPGDRSA 679  
Db 699 -BEGGPKMGGDHQA---DEAGKDKETGDTGLAGSQEHDPQGGSSSPQAGSPTEGEG 754







GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 18, 2003, 15:33:14 ; Search time 19 Seconds  
(without alignments)  
1331.774 Million cell updates/sec

Title: US-09-781-693A-2

Perfect score: 4527

Sequence:

Scoring table: 1 MSRGSGYPHLLNDVKKRGLG.....RLEGDRSGSGOENEDDEE 860

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 120 summaries

Database :

- 1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/iaa/6C\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/iaa/6D\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	429.5	9.5	816	1	US-08-190-802A-54
2	429.5	9.5	816	4	US-08-477-346-54
3	429.5	9.5	816	4	US-08-473-089-54
4	429.5	9.5	816	4	US-08-473-089-54
5	203	4.5	2137	4	US-09-134-001C-4463
6	192.5	4.3	985	5	PCT-US96-03916-6
7	192.5	4.3	985	5	PCT-US96-03916-6
8	189	4.2	1142	3	US-08-993-118-7
9	189	4.2	1142	3	US-08-845-528C-7
10	189	4.2	1142	4	US-08-061-709-2
11	188	4.2	1702	5	PCT-US95-10661A-5
12	188	4.2	1702	5	PCT-US95-10661A-5
13	185.5	4.1	571	4	US-08-961-083-4
14	175.5	3.9	1848	4	US-08-296-791-6
15	175.5	3.9	1848	5	PCT-US95-10661A-6
16	171.5	3.8	1964	2	US-08-790-912-3
17	171.5	3.8	2052	2	US-08-790-912-3
18	169	3.7	664	1	US-08-421-661-6
19	165.5	3.7	1187	1	US-08-320-559-28
20	165.5	3.7	1187	3	PCT-US94-04496-28
21	165.5	3.7	1187	5	PCT-US94-04496-28
22	165.5	3.7	1210	1	US-08-320-559-26
23	165.5	3.7	1210	3	US-08-320-559-26
24	165.5	3.7	1210	5	PCT-US94-04496-26
25	165.5	3.7	1435	2	US-08-568-459A-4
26	165.5	3.7	1435	2	US-08-487-826B-4
27	165.5	3.7	1435	4	US-09-210-288-4

28	164	3.6	278	4	US-08-961-083-94	Sequence 94, Appl
29	154	3.4	713	1	US-08-190-802A-63	Sequence 63, Appl
30	154	3.4	713	4	US-08-477-346-63	Sequence 63, Appl
31	154	3.4	713	4	US-08-473-089-63	Sequence 63, Appl
32	154	3.4	713	4	US-08-487-072A-63	Sequence 63, Appl
33	154	3.4	713	4	US-09-108-857-3	Sequence 3, Appl
34	154	3.4	1093	5	PCT-US93-03077-1	Sequence 1, Appl
35	153.5	3.4	50	1	US-08-190-802A-202	Sequence 202, App
36	153.5	3.4	50	4	US-08-477-346-202	Sequence 202, App
37	153.5	3.4	50	4	US-08-473-089-202	Sequence 202, App
38	153.5	3.4	50	4	US-08-487-072A-202	Sequence 202, App
39	152.5	3.4	1780	1	US-08-769-309A-5	Sequence 5, Appl
40	152.5	3.4	1780	3	US-08-994-570-5	Sequence 5, Appl
41	150.5	3.3	2843	1	US-07-741-940-2	Sequence 2, Appl
42	150.5	3.3	2843	1	US-08-289-548A-2	Sequence 2, Appl
43	150.5	3.3	2843	1	US-08-452-654-2	Sequence 2, Appl
44	150.5	3.3	2843	2	US-08-370-235A-2	Sequence 2, Appl
45	150.5	3.3	2843	2	US-08-449-731-2	Sequence 2, Appl
46	149.5	3.3	2843	4	US-07-853-913-2	Sequence 2, Appl
47	147.5	3.3	1805	1	US-07-741-940-7	Sequence 7, Appl
48	147.5	3.3	2842	1	US-08-289-548A-7	Sequence 7, Appl
49	147.5	3.3	2842	1	US-08-452-654-7	Sequence 7, Appl
50	147.5	3.3	2842	1	US-08-449-731-7	Sequence 7, Appl
51	147.5	3.3	2843	1	US-08-452-655B-2	Sequence 2, Appl
52	147.5	3.3	2843	1	US-08-452-655B-7	Sequence 2, Appl
53	147.5	3.3	2843	3	US-08-450-582-2	Sequence 2, Appl
54	147.5	3.3	2843	3	US-08-821-355A-7	Sequence 7, Appl
55	147.5	3.3	2973	2	US-09-003-687A-7	Sequence 7, Appl
56	147.5	3.3	2973	2	US-09-136-605-7	Sequence 7, Appl
57	147.5	3.3	369	4	US-08-061-376-5	Sequence 5, Appl
58	147.5	3.2	651	1	US-08-769-309A-17	Sequence 17, Appl
59	147	3.2	651	3	US-08-362-525-22	Sequence 22, Appl
60	147	3.2	894	3	US-08-971-692-15	Sequence 15, Appl
61	147	3.2	894	3	US-09-393-569-2	Sequence 2, Appl
62	147	3.2	1360	4	US-08-325-267A-2	Sequence 2, Appl
63	146	3.2	1537	3	US-08-225-728-2	Sequence 2, Appl
64	146	3.2	933	3	US-09-421-868-2	Sequence 2, Appl
65	145	3.2	933	3	US-09-141-047-8	Sequence 8, Appl
66	145	3.2	688	3	US-09-083-521-7	Sequence 7, Appl
67	143	3.2	1317	3	US-08-557-139-2	Sequence 2, Appl
68	143.5	3.1	1162	2	US-08-728-323A-2	Sequence 2, Appl
69	142	3.1	1162	2	US-08-656-068-2	Sequence 2, Appl
70	141.5	3.1	1162	2	US-09-298-534-2	Sequence 2, Appl
71	141.5	3.1	1162	2	US-09-457-040B-27	Sequence 27, Appl
72	139.5	3.1	1584	4	US-08-656-034-2	Sequence 2, Appl
73	139.5	3.1	1584	4	US-09-620-093A-5	Sequence 5, Appl
74	139.5	3.1	1850	4	US-08-656-034-10	Sequence 10, Appl
75	139	3.1	548	3	US-08-760-615-4	Sequence 4, Appl
76	139	3.1	754	4	US-09-214-564A-2	Sequence 2, Appl
77	139	3.1	1863	2	US-08-603-753D-2	Sequence 2, Appl
78	139	3.1	1863	4	US-08-999-753-2	Sequence 2, Appl
79	139	3.1	1863	4	US-08-986-106-2	Sequence 2, Appl
80	139	3.1	1863	4	US-09-007-678B-49	Sequence 49, Appl
81	139	3.1	340	1	US-08-190-802A-42	Sequence 42, Appl
82	138.5	3.1	340	4	US-08-477-346-42	Sequence 42, Appl
83	138.5	3.1	340	4	US-08-473-089-42	Sequence 42, Appl
84	138.5	3.1	340	4	US-08-487-072A-45	Sequence 45, Appl
85	138.5	3.1	949	1	US-09-134-001C-3510	Sequence 3510, App
86	138.5	3.1	1247	4	US-08-961-083-218	Sequence 218, App
87	138.5	3.1	571	4	US-08-893-852A-1	Sequence 1, Appl
88	138	3.0	571	4	US-09-165-239A-4	Sequence 38, Appl
89	138	3.0	1346	2	US-08-477-346-38	Sequence 38, Appl
90	137.5	3.0	341	1	US-08-635-121-2	Sequence 2, Appl
91	137.5	3.0	341	1	US-08-190-802A-45	Sequence 45, Appl
92	137.5	3.0	341	4	US-08-477-346-45	Sequence 45, Appl
93	137.5	3.0	341	4	US-08-473-089-45	Sequence 45, Appl
94	137.5	3.0	491	4	US-08-487-072A-45	Sequence 45, Appl
95	137	3.0	565	4	US-08-961-083-218	Sequence 218, App
96	136.5	3.0	674	3	US-08-893-852A-1	Sequence 1, Appl
97	136.5	3.0	750	4	US-09-165-239A-4	Sequence 38, Appl
98	136	3.0	340	1	US-08-477-346-38	Sequence 38, Appl
99	136	3.0	340	4	US-08-473-089-38	Sequence 38, Appl
100	136	3.0	340	4	US-08-473-089-38	Sequence 38, Appl

us-09-781-693a-2.ra1

Wed Mar 26 09:38:53 2003

101 136 3.0 340 4 US-08-487-072A-38  
102 136 3.0 900 1 US-08-425-061-20  
103 136 3.0 900 2 US-08-825-886-20  
104 136 3.0 914 1 US-08-425-061-21  
105 136 3.0 914 2 US-08-825-886-21  
106 136 3.0 1075 2 US-08-993-228-19  
107 136 3.0 1202 1 US-08-425-061-22  
108 136 3.0 1202 2 US-08-825-886-22  
109 136 3.0 1363 1 US-08-425-061-23  
110 136 3.0 1363 2 US-08-825-886-23  
111 136 3.0 1852 1 US-08-425-061-24  
112 136 3.0 1852 2 US-08-825-886-24  
113 136 3.0 1863 1 US-08-425-061-16  
114 136 3.0 1863 2 US-08-598-591-2  
115 136 3.0 1863 1 US-08-480-784-2  
116 136 3.0 1863 1 US-08-483-553-2  
117 136 3.0 1863 1 US-08-487-002-2  
118 136 3.0 1863 1 US-08-483-554B-2  
119 136 3.0 1863 1 US-08-798-691-2  
120 136 3.0 1863 1 US-08-798-691-4

ALIGNMENTS

RESULT 1  
US-08-190-802A-54  
; Sequence 54, Application US/08190802A  
; Patent No. 5519003  
; GENERAL INFORMATION:  
; APPLICANT: Mochly-Rosen, Daria  
; APPLICANT: Ron, Dorit  
; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses  
; TITLE OF INVENTION: Thereof  
; NUMBER OF SEQUENCES: 265  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dehlinger & Associates  
; STREET: P.O. Box 60850  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306-0850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/190,802A  
; FILING DATE: 01-FEB-1994  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fabian, Gary R. 33,875  
; REGISTRATION NUMBER: 8600-0139  
; REFERENCE/DOCKET NUMBER: 8600-0139  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 324-0880  
; TELEFAX: (415) 324-0960  
; INFORMATION FOR SEQ ID NO: 54:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 816 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; HYPOTHEICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: MUS MUSCULUS PROTEIN, Fig. 37  
; US-08-190-802A-54

Query Match 9.5%; Score 429.5; DB 1; Length 816;  
Best Local Similarity 18.7%; Pred. No. 2.6e-25;  
Matches 164; Conservative 76; Mismatches 173; Indels 465; Gaps 18;

QY 8 PHLLMDV-----RKSLGLEDPRLRSRYLGRREFIQRKLEATLNVDGCVNTICWNTDG 63  
Db 377 PRPRNWLVSALRDLG--SSGRFVYACGARLFVQRFSLHFVFGHSGCVNTVHFNOHG 434  
QY 64 EYILSGSDTKLIVTSNPSYRKVLTTIRSGHRANITSAKFLPCTNDKQIVSCSGDGVIFYT 123  
Db 435 TLLASGDDLVVWMLKRSVNLFDGSHKNNILOAKFLPNCNDAILAMCGRDGQVRVA 494  
QY 124 NVEQDAETNRQCFCHYGTVEIMTVDPYFLSCGEDGTGTFVWFTRIKTSCTKEDCK 183  
Db 495 QLSAVAGTHTKLVKGGASHRLGLEDPSPFRFLTSCGDAVVENIDLRQHPASK----- 550  
QY 184 DDTLLI-----NCRRAATSAICPPPIPYLAVGSDSSVRIYDRMLGTRATGNAGRTTG 239  
Db 551 --LLVKGDKKVGVLTVFVNPANVYQFVAGGQDFMRYIDQKIDENV-----NNG 600  
QY 240 MVARFIPSHL--NKKSCRYTSLCYSDGQELVSYSSDIYFLPKDDTARELKTPTSAEE 297  
Db 601 VLKFCFPHLLSSDYPAHITSLMYSDGTGTEILLASYNDEDIYF----- 643  
QY 298 RRELQPPVKKRLRLRGDWSDTGPRAPSESERERDGEQSPNYSMLQRMDSMLSRWFEEAS 357  
Db 644 ----- 643  
QY 358 EVAOSNRGRGRPRGTSQSDISTLTPVSPDLEVSETAMEVDTPEQFLQPTSTSTM 417  
Db 644 ----- 643  
QY 418 SAQAHSTSPSPHSTPLSSPDSEORQSVESAGHHTHQSDNNNEKLSKPKGTGPVL 477  
Db 644 ----- 643  
QY 478 SLHYSTEGTITIKLNTDEWSSIASSSRGSHCKSEGESEFVPSQSVQPEGDSST 537  
Db 644 ----- 643  
QY 538 KAPEESSEDTVKYQGVSAENPVENHINITOSDKFTAKPLDSNGERNDLDRSCGVE 597  
Db 644 ----- 643  
QY 598 ESASSEKAKEPETSQSTESATNENNTNPEPQFQTEATGSAHEETSTRSALQDTHDS 657  
Db 644 ----- 648  
QY 658 DDDPVLIPGARYRAGPDRRSARVARIQEFFRRKEREKEMEELDTLNIRPLVKMYKGRH 717  
Db 649 D-----GAQY-----AKR-----YKGRH 661  
QY 718 NSRTMIKEANFWG--ANFVMSGDCGHIFIDRHTAHLMLLEAD-NHVVNCLQHPDPDP 774  
Db 662 NNST-VKGVYVYGRSEFVMSGDCGHIFWEKSSQIVQFLEADEGGTINCIDSHPLP 720  
QY 775 ILASSGIDYDIKIWSPLEESRIFNRKLA--DEVITRNMELMEE--TRNTITVPAS---FM 827  
Db 721 VLASSGLDHEVKIWSPIAEP---SKLAGLKNVINKLKRNTFTLRHTSLFNNSMLCEL 777  
QY 828 LRMLASLNH-----IRADRLGDRSGSGQENENEDE 859  
Db 778 MSHVTQSNYGRSWRGIRINAGGGDFSDSSSSSEETNOE 815

RESULT 2  
US-08-477-346-54  
; Sequence 54, Application US/08477346  
; Patent No. 6262023  
; GENERAL INFORMATION:  
; APPLICANT: Mochly-Rosen, Daria  
; APPLICANT: Ron, Dorit  
; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses  
; TITLE OF INVENTION: Thereof  
; NUMBER OF SEQUENCES: 265  
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Morrison & Foerster  
 STREET: 2000 Pennsylvania Avenue, NW  
 CITY: Washington  
 STATE: DC  
 COUNTRY: USA  
 ZIP: 20006-1812  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 FILING DATE: 07-JUN-1995  
 PRIOR APPLICATION NUMBER: 514  
 CLASSIFICATION: 514  
 APPLICATION DATA:  
 FILING DATE: 08/487,072  
 ATTORNEY/AGENT INFORMATION:  
 NAME: MURASHIGE, KATE H.  
 REGISTRATION NUMBER: 29,959  
 REFERENCE/DOCKET NUMBER: 2550-0025.20  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202) 887-1500  
 TELEFAX: (202) 887-0763  
 INFORMATION FOR SEQ ID NO: 54:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 816 amino acids  
 TYPE: amino acid  
 TOPOLOGY: unknown  
 MOLECULE TYPE: protein  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 INDIVIDUAL ISOLATE: MUS MUSCULUS PROTEIN, Fig. 37  
 US-08-477-346-54

Query Match 9.5%; Score 429.5; DB 4; Length 816;  
 Best Local Similarity 18.7%; Pred. No. 2.6e-25;  
 Matches 164; Conservative 76; Mismatches 173; Indels 465; Gaps 18;  
 QY 8 PHLLDWD---RRSLGDEPRLRSRYLGRREFIQRLLKLEATLVNHDGCVNTICWDTG 53  
 Db 377 PRPNVLSALDRDLG--SSGRFVYEACGARLFVQRFSLHFVPGHSGCVNTHFNQHG 434  
 QY 64 EYILSGDDTKLVISNPYSRKVLTTIRSHRANIPSAFLPCTNDKQIVSCGSGVIFYT 123  
 Db 435 TLLASGSDDLKVIYVWDLKRSVLNFDGSHNNILQAKFLPNCNDAILAMCGRDGQVRA 494  
 QY 124 NVEQDAETNROCFTCHYGTTEYIMVPNDPYTFELSCGEDGTWRVEDTRIKTSCTKEDCK 183  
 Db 495 QLSAVAGTHMTKRLVKGHSHRLGLEPDSFRLTSGEDAVVENVNIDLRQHPASK---- 550  
 QY 184 DDILI---NCRRAATSVACPPPIPYILAVGCCSDSSVRIYDRMLGTRATGNVAGRTTG 239  
 Db 551 --LLVIGDKGKVLGYTVFNPNVYQFVAGGQQQFRIYDORKIDENV-----NNG 600  
 QY 240 MVARFIPSHL--NNKSRVTSCLYSEGGQILYSYSDIYLLFDPKDDTARELKTPSAE 297  
 Db 601 VLKFCFPHLLSDYPAPHITSLMYSYDGTILASYNDEDIYF----- 643  
 QY 298 RREELQPPVKRLRLRGDWSDTGPRAPESERERDEQSPNVLQMORMSDMLSRWFEEAS 357  
 Db 644 ----- 643  
 QY 358 EVAQNRGRSRPGTSGSDISTLTPVSPDLVSEVMEVDTPAQFLQPTSTTM 417  
 Db 644 ----- 643  
 QY 418 SAQAHSTSPTEPHSTPLLSPPDSQSRQSVASGHHTHQSDNNNEKLSPKPGTGPVL 477  
 Db 644 ----- 643

QY 478 SLHYSTEGTTTSTIKLNFDEWSSIASSSRGIGSHCKSGEQEESFVPOSSVQPEGDSET 537  
 Db 644 ----- 643  
 QY 538 KAPEESSEDVTKYQGVSAENPVENHINITQSDKFTAKPLDSNGERNDLNLDRSCGYPE 597  
 Db 644 ----- 643  
 QY 598 ESASSEKAKEPETSQDTSTESATINNTNPPQFQTEATGSAHEETSRDALQDITDS 657  
 Db 644 ----- 643  
 QY 658 DDDPVLIPGARYRAGPGDRSAVARIQEFFRRRKERKEMEELDTLNIIRRLVLMVYKGRH 717  
 Db 649 D-----GAQY-----AKR-----YKGRH 661  
 QY 718 NSRTWIKANFWG--ANFVMSGSDCGHIFWDRHTAEHMLLEAD-NHVVNCLOPHFPDP 774  
 Db 662 NNST-VKGVYFYGPRSEFVMSGSDCGHIFWEXSSCOIVQFLEADBGTTINCIDSHPLYP 720  
 QY 775 ILASSGIDYDIKIWSPLEESRIENRKLAA--DEVITRNLML--TRNTITVPAS---FM 827  
 Db 721 VLASSGLDHEVKIWSPIAEP---SKLAGLKNYKINKLKRDNFTLRHTSLFNNSMLCFL 777  
 QY 828 LRMLASLNH-----TRADRLEGDRSGSGQENENEDE 859  
 Db 778 MSHVTQSNYGRSWRGIRINAGGDFSDSSSSSEETNOE 815

RESULT 3  
 US-08-473-089-54  
 ; Sequence 54, Application US/08473089  
 ; Patent No. 6342368  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Mochly-Rosen, Daria  
 ; APPLICANT: Ron, Dorit  
 ; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses  
 ; TITLE OF INVENTION: Theroof  
 ; NUMBER OF SEQUENCES: 265  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Morrison & Foerster  
 ; STREET: 2000 Pennsylvania Avenue, NW  
 ; CITY: Washington  
 ; STATE: DC  
 ; COUNTRY: USA  
 ; ZIP: 20006-1812  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; FILING DATE: 07-JUN-1995  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: MURASHIGE, KATE H.  
 ; REGISTRATION NUMBER: 29,959  
 ; REFERENCE/DOCKET NUMBER: 2550-0025.22  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (202) 887-1500  
 ; TELEFAX: (202) 887-0763  
 ; INFORMATION FOR SEQ ID NO: 54:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 816 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: protein  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; ORIGINAL SOURCE:  
 ; INDIVIDUAL ISOLATE: MUS MUSCULUS PROTEIN, Fig. 37  
 ; US-08-473-089-54



QY 478 SLHYSTEGTSTIKLNFTDEWSSIASRRGIGSHCKSEGOEESFVPOSSVQPPGCDSET 537  
 Db 644 -----  
 QY 538 KAPESSBDVTKYQGYGAENPVENHINITQSDFTAKPLDNSGERNDLNRSCGVPE 597  
 Db 644 -----  
 QY 598 ESASSEKAKEPETSQDOTSTESATNENNTNPQFQTEATGSAHEETSTRDSALQDTDS 657  
 Db 644 -----  
 QY 658 DDDPVLIPGARYRAGPCDRBSAVARIQEFFRRKEREKEMBELDTLNRPLVWVYKGR 717  
 Db 649 D-----GAQY-----AKR-----YKGR 661  
 QY 718 NSTMIKEANFWG--ANFVMSGDCGHIFWDRHTAEHLMLLEAD-NHVVNCLQPPFDP 774  
 Db 662 NNST-VKGVYFYGRSEFVMSGDCGHIFWEKSSQIVQFLEADGGTINCIDSHPLP 720  
 QY 775 ILASSGIDYDIKTWSPLESRIENKLA--DEVITRNEMLLE--TRNTITVPAS---FM 827  
 Db 721 VLASSGIDHEVKVSPIDAP---SKLAGLKNVINKLKNRDNFTLRHTSLFNMSMLCFL 777  
 QY 828 LRLASLNH-----IRADRLGDRSGSGQENENEDE 859  
 Db 778 MSHVTQSNYGRSWRGLRINAGGDFSDSSSEETNQE 815

## RESULT 5

US-09-134-001C-4463  
 ; Sequence 4463, Application US/09134001C  
 ; Patent No. 6380370  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn Doucette-Stamm et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
 ; FILE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: GTC-007  
 ; CURRENT APPLICATION NUMBER: US/09/134.001C  
 ; CURRENT FILING DATE: 1998-08-13  
 ; PRIOR APPLICATION NUMBER: US 60/064,964  
 ; PRIOR FILING DATE: 1997-11-08  
 ; PRIOR APPLICATION NUMBER: US 60/055,779  
 ; PRIOR FILING DATE: 1997-08-14  
 ; NUMBER OF SEQ ID NOS: 5674  
 ; SEQ ID NO 4463  
 ; LENGTH: 2137  
 ; TYPE: PRT  
 ; ORGANISM: Staphylococcus epidermidis  
 US-09-134-001C-4463

Query Match 4.5%; Score 203; DB 4; Length 2137;  
 Best Local Similarity 25.5%; Pred. No. 8e-07;  
 Matches 89; Conservative 52; Mismatches 192; Indels 16; Gaps 9;  
 QY 317 SDTGPRARPERERDEGQSPNVSLMQRMSDMLSRWFEEASEVAQSNRGRSRPRGGTS 376  
 Db 1248 SDSASTSVSDSE-SASTSISESLSTSVSDTS---TSTDASTSTSESDTSSESL 1303  
 QY 377 QSDISTL---PTVPSPDLEVETAMEVDTPAEQFLOPSTSSMTSAQAHSSTSPESPHS 433  
 Db 1304 SESISTSVSDTSASTSDASTSTSESDASTSLSGSTSLSLSDSTSTSDSAS 1360  
 QY 434 TPLLSPDSEQSQVSEAGHHHTHSDNNNEKLSKPKGTGEPVLSLHYSTEGTTSTIKL 493  
 Db 1361 TS-TSESDSE-RASVSLSGSTSLSDSTSTSDASTSTSV-SDSNSASTSLSGSLST 1417  
 QY 494 NPTDEWSSIASRRGIGSHCKSEGOEESFVPOSSVQPPGDSKETAPESSESDVTKYQEG 553  
 Db 1418 SVSDSTSTSDSAS-ASTSESDSERASTSLSGSTSTSLSDSTSTSDASTSTSVSES 1476  
 QY 554 VSAENPVENHINITQSDFTAKPLDNSGERNDLNRSCGVPEESASS--EKAKEPETS 611

Db 1477 NSTSTSESLSLSTVSDSTSTSDASTSTSVSDSASTSSSESVSTSDSESTSTSS 1536  
 QY 612 DOTSTESATNNTNPQFQTEATGSAHEETSTRDSALQDTDDDD 660  
 Db 1537 DSASTSTSVSESNSTSTSLSGSTSTSVSDSTSTSTSDSASASTSESDSD 1585

RESULT 6  
 PCT-US96-03916-6  
 ; Sequence 6, Application PC/TUS9603916  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wild, Martha A.  
 ; APPLICANT: Cochran, Mark D.  
 ; TITLE OF INVENTION: RECOMBINANT INFECTIOUS LARYNGOTRACHEITIS VIRUS  
 ; TITLE OF INVENTION: AND USES THEREOF  
 ; NUMBER OF SEQUENCES: 72  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Cooper & Dunham LLP  
 ; STREET: 1185 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: U.S.A.  
 ; ZIP: 10036  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US96/03916  
 ; FILING DATE: 23-MAR-1995  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/126,597  
 ; FILING DATE: 24-SEP-1993  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: White, John P.  
 ; REGISTRATION NUMBER: 28,678  
 ; REFERENCE/DOCKET NUMBER: 39116-A  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 278-0400  
 ; TELEFAX: (212) 391-0525  
 ; INFORMATION FOR SEQ ID NO: 6:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 985 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 PCT-US96-03916-6

Query Match 4.3%; Score 192.5; DB 5; Length 985;  
 Best Local Similarity 20.7%; Pred. No. 1.6e-06;  
 Matches 173; Conservative 101; Mismatches 290; Indels 273; Gaps 41;  
 QY 172 RIKTSCTKEDCKDILLINCRRATSAICPPIPYLVAGSDSSVYDR-RMLGTRATG 230  
 Db 146 RLETPCDENFLQNEPTWGSKR---WLGPPSY-----VRDNDVAVLTAKAIGECYSN 195  
 QY 231 NVAGRGTGTVARFI--PSHLNN-----KSCRVTSLCYSDGQELVYSYSDYILF 280  
 Db 196 SAAQTGLTSLNFTFFYPKPRIVNTWTGGPSPRTVTVSSRENGQPVLRNVSDGLVKY 255  
 QY 281 DKDD-----TA-RELKTPSAERREELQPPVKRLRLRGDWS 317  
 Db 256 TPDIDGRAMINVIANYSPADSGSVLAFTAFRCGLPSAQLH-----RIDMS 302  
 QY 318 DTGPRARPERERDEGQSPNVSLMQRMSDMLSRWFEEASEVAQSNRGRSRPRGGTSQ 377  
 Db 303 GTEP---FGTETTFD-----CQKM-----IETPYRALGSNVPRDDSTR 337  
 QY 378 SDISTLTPVSPDLE--VSETAMEVDTPAEQFLOPSTSSMTS--AAHSTSSPTESP-- 431





Patent No. 5997872  
 GENERAL INFORMATION:  
 APPLICANT: LUCAS, Sophie;  
 APPLICANT: DE SMET, Charles;  
 APPLICANT: BOON-FALLEUR, Thierry  
 TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR TUMOR  
 TITLE OF INVENTION: REJECTION ANTIGEN PRECURSOR MAGP-C1 AND USES  
 NUMBER OF SEQUENCES: 14  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Felfe & Lynch  
 STREET: 805 Third Avenue  
 CITY: New York City  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10022  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage  
 COMPUTER: IBM PS/2  
 OPERATING SYSTEM: PC-DOS  
 SOFTWARE: Wordperfect  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/993,118  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/08/845,528  
 FILING DATE: April 25, 1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Mary Anne Schofield  
 REGISTRATION NUMBER: 36,669  
 REFERENCE/DOCKET NUMBER: LUD 5455  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 688-9200  
 TELEFAX: (212) 838-3884  
 INFORMATION FOR SEQ ID NO: 7:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1142  
 TYPE: amino acids  
 STRANDEDNESS: single-stranded  
 TOPOLOGY: linear  
 US-08-993-118-7

Query Match 4.2%; Score 189; DB 2; Length 1142;  
 Best Local Similarity 21.0%; Pred. No. 3.9e-06;  
 Matches 138; Conservative 96; Mismatches 282; Indels 142; Gaps 29;  
 QY 271 SYSDYIYLPDKDDTARE-----LKTPSAERREELRQPPVKRLRLGWDSDT-G 320  
 DB 486 SSSSTLLSLFQSSPECTQSTFEGFPQSLQIPQSPPEGENTHS-PLQIVPSLPWEDESL 544  
 QY 321 PRAPESERERDGEQSPNV---SLMQMSDMLSRWFEEA--SEVAQSNRGRSRPRGGT 375  
 DB 545 PHYFPQSPQGEDSLSPHYFPQSPQGEDSLSPHYFPQSPQGEDSLSPHYFPQSPQGED 604  
 QY 376 SOSDITLTVPSPDLEVSETAMEVDTPAQFLPQSTSTMSAQAHSTSSPESPHSTP 435  
 DB 605 SMSPL-----YFPQSP-LQGEFQSSLSQSPVSTCSSSTPSLSQSPFESSQSPPEGVQSP 659  
 QY 436 LLSPPDSQEQRSVEASGHHTHQSDNNNEKLSG-----KPGTGEPLV-SLHY--STEGTT 487  
 DB 660 -LHSPQSPMECHMSQSPLOSPEASPEGEDSLPLQIPQSPLEGEDSLSLHPQSPPEWE 718  
 QY 488 TSTIKLNT-----DEWSSIASSTRGIGSHCKSGQESFVQSSVOPPEGDSST--KA 539  
 DB 719 DSLSPHPQPPQGEDPQSSLSQSPVSTCSSSTSLQSPF-PESQSPPEGPAQSPQR 777  
 QY 540 P-----EESDVTYKQGVGVS-----AENPVENHINITQSDKFTAKPLDINSGE 583  
 DB 778 PVSFFSTLASLQSHESQSPPEGPAQSPLOSQSPVSSPSSSTSSLSQSPVSS----- 833  
 QY 584 RNDLNRSCGVPPEASASSEKAKPEPDSQSTESATNNTNPEPQFTATGSPSAHEE 643

DB 834 -----FPSSTSSLSKSSPESPLQSPVIFSSTSLSP-----FSESSSP-VDEV 878  
 QY 644 TSTRDSALQDITDSDDDPVLIPGARYRAGPGRDRSAVAR-----582  
 DB 879 TSSDITLESLSLTDSESLIESEPLFTVTLDEKVELARFLLKLYQVQKQPTKAEMLTNV 938  
 QY 683 -----IQEFFRRRKRKE-----MEELD-----TLNIRPLVKMYKGRHNSR 720  
 DB 939 ISRYTGVPVIFRKAREFIELFGISLREVDPDSDSYFVNTIDLTSEGCLSDQSGMSQNR 998  
 QY 721 TWIKEANFWANFVMSGDCGHIFWD-----RHTAEHLMLLEADNHVNC-LQPH-- 770  
 DB 999 LLILILSI-----IFIKTYAEEVWDVLSGIGYVGRGEHFAFGEPRELLTKVWQEHYL 1054  
 QY 771 PFDPIASSGIDYDIKWSPLEESRIFNRKLADDEVITRNELMLEETRNT--ITVPASF 836  
 DB 1055 EYREVPNSPPRYEF-LWGPRAHSEVIRKVV-----FLAMLKNTVPIITFPSSY 1103  
 RESULT 9  
 US-08-845-528c-7  
 ; Sequence 7, Application US/088455528c  
 ; Patent No. 6027924  
 ; GENERAL INFORMATION:  
 ; APPLICANT: LUCAS, Sophie;  
 ; APPLICANT: DE SMET, Charles;  
 ; APPLICANT: BOON-FALLEUR, Thierry  
 ; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR TUMOR  
 ; TITLE OF INVENTION: REJECTION ANTIGEN PRECURSOR MAGP-C1 AND USES  
 ; NUMBER OF SEQUENCES: 14  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Felfe & Lynch  
 ; STREET: 805 Third Avenue  
 ; CITY: New York City  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10022  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage  
 ; COMPUTER: IBM PS/2  
 ; OPERATING SYSTEM: PC-DOS  
 ; SOFTWARE: Wordperfect  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/845,528c  
 ; FILING DATE: April 25, 1997  
 ; CLASSIFICATION: 4335  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Mary Anne Schofield  
 ; REGISTRATION NUMBER: 36,669  
 ; REFERENCE/DOCKET NUMBER: LUD 5455  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 688-9200  
 ; TELEFAX: (212) 838-3884  
 ; INFORMATION FOR SEQ ID NO: 7:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1142  
 ; TYPE: amino acids  
 ; STRANDEDNESS: single-stranded  
 ; TOPOLOGY: linear  
 ; US-08-845-528C-7  
 Query Match 4.2%; Score 189; DB 3; Length 1142;  
 Best Local Similarity 21.0%; Pred. No. 3.9e-06;  
 Matches 138; Conservative 96; Mismatches 282; Indels 142; Gaps 29;  
 QY 271 SYSDYIYLPDKDDTARE-----LKTPSAERREELRQPPVKRLRLGWDSDT-G 320  
 DB 486 SSSSTLLSLFQSSPECTQSTFEGFPQSLQIPQSPPEGENTHS-PLQIVPSLPWEDESL 544  
 QY 321 PRAPESERERDGEQSPNV---SLMQMSDMLSRWFEEA--SEVAQSNRGRSRPRGGT 375  
 DB 545 PHYFPQSPQGEDSLSPHYFPQSPQGEDSLSPHYFPQSPQGEDSLSPHYFPQSPQGED 604

```

376 QY SQSDISTLTPVSSPDLEVSAMEVDTPAEQFLQFSTSTMSAQAHSTSSPESHPT 435
376 QY SQSDISTLTPVSSPDLEVSAMEVDTPAEQFLQFSTSTMSAQAHSTSSPESHPT 435
605 Db SMSPL-----YFPQSP-LQGEFQSSLOSPVSCSSSTPSSLPQSPFESSQSPGVPQSP 659
605 Db SMSPL-----YFPQSP-LQGEFQSSLOSPVSCSSSTPSSLPQSPFESSQSPGVPQSP 659
436 QY LLSPPDSEORQSVASGHHTHQSDNNNEKLS-----KPGTGPVL-SLHY--STEGTT 487
436 QY LLSPPDSEORQSVASGHHTHQSDNNNEKLS-----KPGTGPVL-SLHY--STEGTT 487
660 Db -LHSPQSPGEMHQSPLQSPESAPAGEDSLPLQIPQSPLEGEDSLSLHFPQSPPEWE 718
660 Db -LHSPQSPGEMHQSPLQSPESAPAGEDSLPLQIPQSPLEGEDSLSLHFPQSPPEWE 718
488 QY TSTIKLNT-----DEWSSIASRRGIGSHCKSEGEESFVQSSVQPPGDSST--KA 539
488 QY TSTIKLNT-----DEWSSIASRRGIGSHCKSEGEESFVQSSVQPPGDSST--KA 539
719 Db DLSPLHFPQPPQGEDFQSSLOSPVSCSSSTSLSPQSF-PESQSPGPAQSPLOQR 777
719 Db DLSPLHFPQPPQGEDFQSSLOSPVSCSSSTSLSPQSF-PESQSPGPAQSPLOQR 777
540 QY P-----ESESDEVTKYQEGVS---AENPVENHINITQSDKFTAKPLDSSNGE 583
540 QY P-----ESESDEVTKYQEGVS---AENPVENHINITQSDKFTAKPLDSSNGE 583
778 Db PVSSFFSYTLASLLQSSHESPPQSPGPAQSPLOSPVSSFPSSSTSSLSQSPVSS-- 833
778 Db PVSSFFSYTLASLLQSSHESPPQSPGPAQSPLOSPVSSFPSSSTSSLSQSPVSS-- 833
584 QY RNDNLDRSCGVPESASKEKAPETSDQTSATNENNTNPEQFOTATGSAHEE 643
584 QY RNDNLDRSCGVPESASKEKAPETSDQTSATNENNTNPEQFOTATGSAHEE 643
834 Db -----FPSSTSSLSKSPSPLOSPVIFSSSTSLSP---FSESSSP-VDEY 878
834 Db -----FPSSTSSLSKSPSPLOSPVIFSSSTSLSP---FSESSSP-VDEY 878
644 QY TSTRSALQDSDDDPVLIPGARYRAGPDRRSARV----- 682
644 QY TSTRSALQDSDDDPVLIPGARYRAGPDRRSARV----- 682
879 Db TSSDYLLESDSLDSSESLIESEPLTYTLDEKVDLAREFLLLKYQVKQPIKAEMLTNV 938
879 Db TSSDYLLESDSLDSSESLIESEPLTYTLDEKVDLAREFLLLKYQVKQPIKAEMLTNV 938
683 QY -----IQEFFRRKKE-----MEELD-----TLNIRPLVMKYKGHRNSR 720
683 QY -----IQEFFRRKKE-----MEELD-----TLNIRPLVMKYKGHRNSR 720
939 Db ISRYTGVPVIFRKAREFIEILFGLISREVDPDSDSVVFNLTDLTSEGCLSDGQMSQNR 998
939 Db ISRYTGVPVIFRKAREFIEILFGLISREVDPDSDSVVFNLTDLTSEGCLSDGQMSQNR 998
721 QY TMKEANFWANFVMSGDCGHIFWD-----RHTAHLMLLEADNHVNC-LQPH-- 770
721 QY TMKEANFWANFVMSGDCGHIFWD-----RHTAHLMLLEADNHVNC-LQPH-- 770
999 Db LLILILSI-----IFKGTAYASEEIVDVLGIGVRAGREHFAFGEPRELLTKVWQEHYL 1054
999 Db LLILILSI-----IFKGTAYASEEIVDVLGIGVRAGREHFAFGEPRELLTKVWQEHYL 1054
771 QY PFDPIASSGIDYDIKINSPLESRIFNRKLADIVITRNLMLLETRNT--ITVPASF 826
771 QY PFDPIASSGIDYDIKINSPLESRIFNRKLADIVITRNLMLLETRNT--ITVPASF 826
1055 Db EYREVPNSPPRYEF-LMGPRAHSEVIRKKVVE-----FLAMKNTVPTITPSSY 1103
1055 Db EYREVPNSPPRYEF-LMGPRAHSEVIRKKVVE-----FLAMKNTVPTITPSSY 1103

RESULT 11
US-08-296-791-5
; Sequence 5, Application US/08296791
; Patent No. 6245337
; GENERAL INFORMATION:
; APPLICANT: St. Gene III, Joseph W.
; APPLICANT: Falkow, Stanley
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; TITLE OF INVENTION: Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/296,791
; FILING DATE: 25-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Treccartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299

RESULT 10
US-09-061-709-2
; Sequence 2, Application US/09061709B
; Patent No. 6297364
; GENERAL INFORMATION:
; APPLICANT: Chen, Yao-Tseng
; APPLICANT: Gurs, Ali
; APPLICANT: Tsang, Solam
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Jager, Elke
; APPLICANT: Knuth, Alexander
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated
; TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof
; FILE REFERENCE: LUD 5538
; CURRENT APPLICATION NUMBER: US/09/061,709B
; CURRENT FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 2
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; US-09-061-709-2

Query Match 4.2%; Score 189; DB 4; Length 1142;
Best Local Similarity 21.0%; Pred. No. 3.9e-06;
Matches 138; Conservative 96; Mismatches 282; Indels 142; Gaps 29;

QY 271 SYSSDYILFDPKDDTARE-----LKTPSAERREERLEPPVYKRLRLRGDWSDT-G 320
Db 486 SSSSTLLSQSSPECTQSTFEGFPQSPLOIPQSPGENTHS-PLQIVPSLPEWEDSL 544
QY 321 PRAPERERERDQSPNV-----SLMQMSDMLSEWFEA---SEVAQSNRGRSRPRGGT 375

```



us-09-781-693a-2.ra1

wed Mar 26 09:38:53 2003

Db 110 SDSQOLAEHKKLEKKEKISPKKKTGVTNLNPODEVLSQOLNKPPELLYREETMETID 169  
 QY 272 Y8S-----DYIYLFDPKDDATRELKTPSAERREELRQP 305  
 Db 170 FOEIEQENPLAAGTVRVKQEGKGLKKKVEIRIFSVNKEEVSREIVSTST--TA 221  
 QY 306 PVKRLRLRGDMSDTGPRAPESERERDGEQS-----PNVSLMQMSDMLSRWFEEASEVA 360  
 Db 222 PSPRIKVEKGGKTKQVKEQPETGVEHKDVQSGAIVRPAIQ-----PELPEAV 268  
 QY 361 QSNRGRGRSPRGTSOSDSTLPTV-----PSSPDLEVSETA----- 398  
 Db 269 VSDKGEVOP-----TLPEAVTVTKGETVQPESDPTVWSDKGPEQVAPLPEY 318  
 QY 399 -----NEVDTPARQFQTPSTSTMSAQHSTSPSTPHSTLLSSPDSEORQSVASG 452  
 Db 319 KGNTEQVKEPPEVKEQGEKPEEV-----PVKPTETPVNPNNEGTEGTSIQEAE 371  
 QY 453 HHTH--HOSDNNNEKLPKPGTGEVLSHYSTEGTTTSTIKLFTDWSIASSSRGIG 510  
 Db 372 NPVQPAEESTTNSEKVP-----DTSKNTGEVSSNPSDSTTSVGSSEKPEH 418  
 QY 511 SHCKSEGOEESFVQSSVQPPGDSKAPESSESDVTKYQEGVSAENPVENHINITQSD 570  
 Db 419 NDSKNENSEKT-VEEVVNPNEGTVEGTSNQETKPVQPAEE-----TOTN 463  
 QY 571 KFTAKPLDSNGERDNLDRSCGVPEESASSEK---AKEPETSDQSTESATNENNTNP 627  
 Db 464 --SGKIANENTGEVSNKPSDSRPPV-EESNQPEKNGTATKPNESGNTTSEN---GOTEP 516  
 QY 628 EPQFQTEATGPSAHEETSTRDSALQDTSDDDDP 661  
 Db 517 EPSNGNSTEDVSTESNTSNGNEIKQENELDP 550

RESULT 14

US-08-296-791-6  
 ; Sequence 6, Application US/08296791  
 ; Patent No. 6245337  
 ; GENERAL INFORMATION:  
 ; APPLICANT: St. Geme III, Joseph W.  
 ; APPLICANT: Falkow, Stanley  
 ; TITLE OF INVENTION: Haemophilus Adherence and Penetration  
 ; TITLE OF INVENTION: Protein  
 ; NUMBER OF SEQUENCES: 9  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESS: Flehr, Hobbach, Test, Albritton & Herbert  
 ; STREET: 4 Embarcadero Center, Suite 3400  
 ; CITY: San Francisco  
 ; STATE: California  
 ; COUNTRY: United States  
 ; ZIP: 94111-4187  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent in Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/296,791  
 ; FILING DATE: 25-AUG-1994  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Trecartin, Richard F.  
 ; REGISTRATION NUMBER: 31,801  
 ; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (415) 781-1989  
 ; TELEFAX: (415) 398-3249  
 ; TELEX: 910 277299  
 ; INFORMATION FOR SEQ ID NO: 6:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1848 amino acids

QY 459 SDNNNEKLSKPGTGEVLSHYSTEGTTTSTIKLFTDWSIASSSRGIGSHCKSEGO 518  
 Db 1212 TENT-----TDQP-----TEREKTAKVETETQPPQVASQA-----SPKQEQ 1249  
 QY 519 EESFVQSSVQPPGDSSTKAPESSESDVTKYQEGVSAENPV-ENHIN-----ITQSDK 571  
 Db 1250 SETVQPAVLESENVPVNNAAEEVQAQLOTQTSATVSTKQAPENSINTGSATAITETAE 1309  
 QY 572 FVAKPLDSNGERDNLDRSCGVPEESASSEKAKEPETSQDSTESATNENNTNPEQOF 631  
 Db 1310 KSDKP-----QTETAATEDASOHRKANTVADNSVANNSESSEPKSRR 1351  
 QY 632 QTEATGPSAHEETSTRDSALQDTSDDDDVLPICARYAGGDRSARVARIOEFFFFRRK 691  
 Db 1352 RRSISQ---QETSABETTAASDTET-----TIADNSKRKSKP-NRRS-----RRS 1392  
 QY 692 ERKEMEELDTLNRRLPVLMVYKGRHSRTMIKEANFWANGFVMSGDCGHIFI---WDR 748  
 Db 1393 VRSE-----PTVYNGSDRSVALRDLTSTNTNAVISDAMAKAQFVALNVCK 1438  
 QY 749 HTAEHLMLLEADN 761  
 Db 1439 AVSQHISQLENN 1451

RESULT 13

US-08-961-083-4  
 ; Sequence 4, Application US/08961083  
 ; Patent No. 6159469  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Choi et. al.  
 ; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
 ; NUMBER OF SEQUENCES: 452  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Human Genome Sciences, Inc.  
 ; STREET: 9410 Key West Avenue  
 ; CITY: Rockville  
 ; STATE: Maryland  
 ; COUNTRY: USA  
 ; ZIP: 20850  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
 ; COMPUTER: Hp Vectra 486/33  
 ; OPERATING SYSTEM: MSDOS version 6.2  
 ; SOFTWARE: ASCII Text  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/961,083  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Brookes, A. Anders  
 ; REGISTRATION NUMBER: 36,373  
 ; REFERENCE/DOCKET NUMBER: PB340P2  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (301) 309-8504  
 ; TELEFAX: (301) 309-8512  
 ; INFORMATION FOR SEQ ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 571 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-961-083-4  
 ; 4.1k: Score 185.5; DB 4; Length 571;  
 ; Query Match  
 ; Best local Similarity 19.3%; Pred. No. 2.6e-06;  
 ; Matches 99; Conservative 80; Mismatches 198; Indels 137; Gaps 17;  
 ; QY 212 SDSVRIYDRMLGTRATGNVAGRGTTGMVARTPSPHNNKSCRVTSLCYSDGQELVLS 271

TYPE: amino acid  
TOPOLOGY: unknown  
US-08-296-791-6

Query Match  
Best Local Similarity 3.9%; Score 175.5; DB 4; Length 1848;  
Matches 106; Conservative 68; Mismatches 218; Indels 99; Gaps 19;  
QY 250 NKKSCRV-----TSLCYSDGQELVSYSSDYIYLFDPKDDTARELKTPTSAEER 299  
Db 920 NKKSNKVVVVKNSATGNTLQVADKTGP-----NHNELTLFDASNATRNLEVTLANGSV 974  
QY 300 EELRQPPVKRLRLRGDWS-----DTGPR---ARPESERERDGEQSPNVSLMORMSDMLSRW 352  
Db 975 D-----RGAWKYKLRNVNGRYDLYNEVEKRNQTVDTTNIITPNDI----- 1015  
QY 353 FEAESEVAOSNRGRSRPRGGTSQSDISTL---PTVPSSPDLEVSETAMEVDTPAEQFLQ 410  
Db 1016 -QADAPSAQSN-----NEEIARVETPVPPAPAPATESAIAEQPETAETAQ 1061  
QY 411 P-----STSTMSQAQHSSTSPSTPSHTPLSSPDSEQSQSVASGHHTHQSDNNNE 464  
Db 1062 PAMEETNTANSTETAPKSDTATQENPSE---SVPSETTEKVAENPPQENETVAKNEQE 1118  
QY 465 KLSKPGTGE-----PVLSLHYSTEGTTSTIKLNFDEWSSIASSSRGIGSHCKSEGO 518  
Db 1119 ATEPTPQNGEVAKEQDPVEANTQNEATQSEK---TEETQTAETKSEPTESVTVSENO 1175  
QY 519 BESFVQSSVOPPEGDSKAPESSESDVTKYQGVSAENPVENHINTQSDKFTAK-PL 577  
Db 1176 PEKTVSQSTEDKVVVEKEAKVE-TEETQKAPQVTSKEPP-----KOAEPAPAEVPT 1227  
QY 578 DSNSEGERNDLNRSCGVPESASSEKAKPETSDQSTESATNENNTNPEQFQTEATG 637  
Db 1228 DTNAEEAQAALQOQPTTVAAEETSPNSKPAEETQOPS-----EKTNAEP-----VT 1274  
QY 638 PSAHEETSTRDSALQDQDSDDDPVL-IPGARYRAGPGRDRSAVARIOEFFRRKREKEM 696  
Db 1275 PVVSENTAQPTTEETAKVEKEKTOEVQVQASQESPKQEQPA-AKPOAQTKPQAPARE 1333  
QY 697 EELDTLNIRRP 707  
Db 1334 NVLTTKNVGEP 1344

RESULT 15  
PCT-US95-10661A-6  
Sequence 6, Application PC/TUS9510661A  
GENERAL INFORMATION:  
APPLICANT: Washington University, et al.  
TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/10661A  
FILING DATE: 16-AUG-1995  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/296,791  
FILING DATE: 25-AUG-1994  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:

NAME: Trecartin, Richard F.  
REGISTRATION NUMBER: 31,801  
REFERENCE/DOCKET NUMBER: FP-59941/RFT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1848 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
PCT-US95-10661A-6

Query Match  
Best Local Similarity 3.9%; Score 175.5; DB 5; Length 1848;  
Matches 106; Conservative 68; Mismatches 218; Indels 99; Gaps 19;  
QY 250 NKKSCRV-----TSLCYSDGQELVSYSSDYIYLFDPKDDTARELKTPTSAEER 299  
Db 920 NKKSNKVVVVKNSATGNTLQVADKTGP-----NHNELTLFDASNATRNLEVTLANGSV 974  
QY 300 EELRQPPVKRLRLRGDWS-----DTGPR---ARPESERERDGEQSPNVSLMORMSDMLSRW 352  
Db 975 D-----RGAWKYKLRNVNGRYDLYNEVEKRNQTVDTTNIITPNDI----- 1015  
QY 353 FEAESEVAOSNRGRSRPRGGTSQSDISTL---PTVPSSPDLEVSETAMEVDTPAEQFLQ 410  
Db 1016 -QADAPSAQSN-----NEEIARVETPVPPAPAPATESAIAEQPETAETAQ 1061  
QY 411 P-----STSTMSQAQHSSTSPSTPSHTPLSSPDSEQSQSVASGHHTHQSDNNNE 464  
Db 1062 PAMEETNTANSTETAPKSDTATQENPSE---SVPSETTEKVAENPPQENETVAKNEQE 1118  
QY 465 KLSKPGTGE-----PVLSLHYSTEGTTSTIKLNFDEWSSIASSSRGIGSHCKSEGO 518  
Db 1119 ATEPTPQNGEVAKEQDPVEANTQNEATQSEK---TEETQTAETKSEPTESVTVSENO 1175  
QY 519 BESFVQSSVOPPEGDSKAPESSESDVTKYQGVSAENPVENHINTQSDKFTAK-PL 577  
Db 1176 PEKTVSQSTEDKVVVEKEAKVE-TEETQKAPQVTSKEPP-----KOAEPAPAEVPT 1227  
QY 578 DSNSEGERNDLNRSCGVPESASSEKAKPETSDQSTESATNENNTNPEQFQTEATG 637  
Db 1228 DTNAEEAQAALQOQPTTVAAEETSPNSKPAEETQOPS-----EKTNAEP-----VT 1274  
QY 638 PSAHEETSTRDSALQDQDSDDDPVL-IPGARYRAGPGRDRSAVARIOEFFRRKREKEM 696  
Db 1275 PVVSENTAQPTTEETAKVEKEKTOEVQVQASQESPKQEQPA-AKPOAQTKPQAPARE 1333  
QY 697 EELDTLNIRRP 707  
Db 1334 NVLTTKNVGEP 1344

RESULT 16  
US-08-790-912-3  
Sequence 3, Application US/08790912  
Patent No. 5976542  
GENERAL INFORMATION:  
APPLICANT: Weiser, Jeffrey N.  
APPLICANT: Plaut, Andrew G.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT  
OF STREPTOCOCCUS PNEUMONIAE INFECTION  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.  
STREET: 1601 Market Street, 36th Floor  
CITY: Philadelphia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19103-2398  
COMPUTER READABLE FORM:

us-09-781-693a-2.rai

Wed Mar 26 09:38:53 2003

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/790,912
  FILING DATE: 29-JAN-1997
  CLASSIFICATION: 424
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 60/026,939
    FILING DATE: 23-SEP-1996
  ATTORNEY/AGENT INFORMATION:
    NAME: Leary, Kathryn
    REGISTRATION NUMBER: 36,317
    REFERENCE/DOCKET NUMBER: 7600-401
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (215) 567-2020
    TELEFAX: (215) 567-2991
    TELEX: 831-494
  INFORMATION FOR SEQ ID NO: 3:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 1964 amino acids
      TYPE: amino acid
      STRANDEDNESS:
      TOPOLOGY: linear
      MOLECULE TYPE: protein
US-08-790-912-3

Query Match      3.8%; Score 171.5; DB 2; Length 1964;
Best Local Similarity 18.1%; Pred. No. 0.00023;
Matches 121; Conservative 101; Mismatches 229; Indels 217; Gaps 24;

QY 111 IVSCSGDVIFVTNVEQDAETNROCQFTCHYGTIYEI-----MTVPND--PYTFLSGCGD 163
DB 139 LVAGMGVLSVHALEN-----HLLNNTDYELTSGEKLPLPKKEISGYTIVGIKE 191
QY 164 G-----TVWFDTRIKTSCTK----- 179
DB 192 GKTTDFEVSNOEKSAAFTKQKVDYNTVPNFVDPSTVQAIQEQTPVSSTKTEVQVV 251
QY 180 EDCKDDILLINCRRAATVAICPPPIYLVAGCSDSVRIYDRMLGTRATGNAGRGTTG 239
DB 252 EKPFSTELINPRKEEQ-----SDSQEQLAEHKNLETKKEEKISPKKGTG 297
QY 240 MVAFIPSHLNKSCRVTSICYSEDGOEILVSYSS----- 274
DB 298 VNTLNPODEVLSGQLNKPELLEETIETKIDFQEEIQENPDLAECTVRVKQEGKLGKV 357
QY 275 DYIYLFDPKDDTARELKTPSAEERBELQPPVKRLRLGDSWDGPRAPESERERDG 333
DB 358 ELVRIFSVNKEEVSREIVSTST-----TAPSPRIVEKGTGKTKQVKEQPTGVYEHKD 409
QY 334 EQS-----PNVSLMQMSDLMSRWFEBASEVAQSNRGRSRPRGGTSQSDISTLPTV-- 386
DB 410 VQSGAIVEPAIQ-----PELPEAVVSDKGEPEVQ-----TLPEAVV 446
QY 387 -----PSSPDLEVSETA-----MEVDTPEAQFLOPSTSSYMSAQ 420
DB 447 TDKGETEVQSPDFTVSDKGEQVAPLBYKGNIEQVKPETPVEKKEQGPKEEV- 505
QY 421 AHSTSSPTSPSTPLSSPDSEQRQSVASGHHTH--HQSDNNNEKLSPKPGTGPVLS 478
DB 506 -----PVKPTTEPTVPNNEGTEGTSIQEAPNPVQPAEESTNSEKVP----- 549
QY 479 LHVSTEGTTTSLKLNFTDEWSIASRGICSHCKSGQESFVPQSSVQPPGDSSTK 538
DB 550 ----DTSSSENTGEVSNPSSDSTTSVGSNKPENHDSKNSEKT--VEEVPVNPNEGTVGT 605
QY 539 APESSEEDVTKYQGVSAENPVENHINITQSDKFTAKPLDLSNGSGERNDLNRSCGVPE 598
DB 606 SNQETKPVQPAEE-----TQTN--SGKIANENTGEVSNKPSDKPPV-EE 648
QY 599 SASSEK---AKEPETSQDTSTESATNE-----NNTNPEQFQTEATGSAHEETSTR 647
DB 649 SNOPEKNGTATKPSNGNTTSENGQTEPEKPEKLELRNVSDIELYSQTNGT-----YRQHVS 704
QY 648 DSALQDTD 655
DB 705 DGIPEND 712

RESULT 17
US-08-790-912-2
; Sequence 2, Application US/08790912
; Patent No. 5976542
; GENERAL INFORMATION:
; APPLICANT: Weiser, Jeffrey N.
; APPLICANT: Plaut, Andrew G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT
; TITLE OF INVENTION: OF STREPTOCOCCUS PNEUMONIAE INFECTION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
; STREET: 1601 Market Street, 36th Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103-2398
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/790,912
; FILING DATE: 29-JAN-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,939
; FILING DATE: 23-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Leary, Kathryn
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: 7600-401
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 567-2020
; TELEFAX: (215) 567-2991
; TELEX: 831-494
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2052 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-790-912-2

Query Match      3.8%; Score 171.5; DB 2; Length 2052;
Best Local Similarity 18.1%; Pred. No. 0.00023;
Matches 121; Conservative 101; Mismatches 229; Indels 217; Gaps 24;

QY 111 IVSCSGDVIFVTNVEQDAETNROCQFTCHYGTIYEI-----MTVPND--PYTFLSGCGD 163
DB 212 LVAGMGVLSVHALEN-----HLLNNTDYELTSGEKLPLPKKEISGYTIVGIKE 264
QY 164 G-----TVWFDTRIKTSCTK----- 179
DB 265 GKTTDFEVSNOEKSAAFTKQKVDYNTVPNFVDPSTVQAIQEQTPVSSTKTEVQVV 324
QY 180 EDCKDDILLINCRRAATVAICPPPIYLVAGCSDSVRIYDRMLGTRATGNAGRGTTG 239
DB 325 EKPFSTELINPRKEEQ-----SDSQEQLAEHKNLETKKEEKISPKKGTG 370
QY 240 MVAFIPSHLNKSCRVTSICYSEDGOEILVSYSS----- 274
DB 371 VNTLNPODEVLSGQLNKPELLEETIETKIDFQEEIQENPDLAECTVRVKQEGKLGKV 430
```

QY	275	DIYILFD-PKODTARELKT	PSAERRERELQPPVKRLRLGDSWSDTGP	RRARPESEREDG	333		
Db	431	EIVRIFSVNKEVSEI	YSTST-----TAPSPRIVEKGT	TKTQVKEQPETGV	482		
QY	334	EQS-----PNYSLMQ	MSDLRWFEEASEVAOSRGRSRG	TSGDSITLPTV-----	386		
Db	483	VOSGAIVEPA	IQ-----PELPEAVVSDKGE	VEVQ-----TILPEAVV	519		
QY	387	-----PSSPDLEY	SETA-----MEVDTPAEQFLQ	PSSTSMQAQ	420		
Db	520	TDKGETEVQ	PSPTVYVSDKGEQ	EVAPLEPYKGNIEQV	KTPETVEKTEQ	578	
QY	421	AHSTSSPT	SHSPHLLSSPDSQ	ROSVEASGHHTH-----HOSDNNEK	LSPRGTEPVL	478	
Db	579	-----PVKTEET	VPNPEGTEGTSIQ	EANVPQAEESTN	SEKVP-----	622	
QY	479	LHYSTEGTT	TTTKLNFTEWSS	SIASSRGCSHGCSQGEESF	YQSSVPPGDS	538	
Db	623	-----DTSS	ENTGEVSNPSD	STTVSGESKNPEH	DSKNSEKT-VEEVPV	678	
QY	539	APRESSE	DTYKQGVSAENPV	ENHINITQSDKFTAKPLD	SNRNDLNLDR	598	
Db	679	SNQET	EKPQPAEE-----TQIN-SG	KIANTENGEV	KNKPSDKPPV-EE	721	
QY	599	SASSEK	---AKRPETSD	OTSTESATNE-----NNTNPEQ	QTEATGPSAHE	647	
Db	722	SNQKNG	ATAPENS	GNNTTSENGQTEPK	KLRLNVSDIEL	YSQINGT-----YRQVSL	777
QY	648	DSALQ	QDT	655			
Db	778	DGIP	ENT	785			

```

RESULT 18
US-08-421-661-6
; Sequence 6, Application US/08421661
; Patent No. 5807993
; GENERAL INFORMATION:
; APPLICANT: French, Cynthia K
; APPLICANT: Yamamoto, Karen K
; APPLICANT: Chow, Phoebe M
; APPLICANT: Alido, Nemesias T
; TITLE OF INVENTION: Antigen Related to Inflammatory Diseases
; NUMBER OF SEQUENCES: 6

```

REFERENCE ADDRESS:  
ADDRESSEE: David J. Oldenkamp  
STREET: 2029 Century Park East, Suite 3800  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.  
ZIP: 90067

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/421,661  
FILING DATE: 08/08/90

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Oldenkamp, David J  
REGISTRATION NUMBER: 29,421  
REFERENCE/DOCKET NUMBER: 109-067  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 310-788-5000  
TELEFAX: 310-277-1297  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 664 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein

US-08-421-661-6

```
Query Match          3.7%; Score 169; DB 1; Length 664;  
Best Local Similarity 19.8%; Pred. NO. 6.4e-05;  
Matches 117; Conservation
```

00, Mismatches	254;	Indels	134;	Gaps	25;
00, Mismatches	254;	Indels	134;	Gaps	25;

QY	321	PRAPESERERDGEQSPNVSL-----MQRSDMLSNWFEEASEVAQNSRGRGRSRPRG	373
Db	43	PSLRAEDE---ENKETAYSTEDDSHHKAESSVLKSK--EESHEQSAEQGSSSQELG	96
QY	374	GTQSODIS-----TLPTVPSPDLLEVSETAMEVDTPAEQ-----FLQPSTSS-TMSAQ	420
Db	97	LKDQEDSDGDLNVNLEYAPTCTGLDIKE---DMSEPOEKLSNTDFLAGVYSFDSNQ	153
QY	421	AHSTTSPTESPSTPLLLSPDSEQRQSVASGHHTHQSDNNNEKLSPRGTPGEPVLSLH	480
Db	154	QESITKRENOQ-----EQPNYSHQLNRSSHQSISOGLRDQGNQEQDPNISNGE-----	201
QY	481	YSTEGTGYSTIKLNFTDWSSIASRRGISGHSCKSEGQESFPVQSSVQPPCEDSETKAP	540
Db	202	-----EEEEKEPVCVTH-NDNQER-----KTLPREHANSKQE	234
QY	541	EESSEDVTKYQEGVSAENPVENHNITQSKFFAKPLDNSGERNDLNLDRSCGVPEESA	600
Db	235	EDNTQ-----SDDLLESDOPTQVSKMQDEFF-----DQCNOQEED--NSNAEMEENA	281
QY	601	SSEKAKEPTSQ-----TSTESATNENNTPNPQFOFATEGPSAHESTSTRDSALQDTD	655
Db	282	SNVNKHIOETEWOSQEGTKGLEAISNHKETEKTVSEALLMEPTDDGGNTTPRNHGVDODG	341
QY	656	DSDDDDPVLIPGARVRAG-----PGDRRSAVA---RIQEFFRRRKERKEWEELDT	701
Db	342	DDCGDDGDTGPRLASDHVFIPSAQFALEARAQSIAYHLKIE-----QREKVHENEN	395
QY	702	LNTRRPLVMVKYKGHRNSTMIKEANFWCANFMVMSGDC-----GHIFWDRHTAEHL	754
Db	396	IGTTECEHQEAQKAABNSSNE-EETSSEGNRVHAVDVSCMSFOCKRGHICADQQGPKHC	454
QY	755	MLEADNHYVNCIQLPHPTDPIILASSGIDYDIKIWSPLEESRIFNRKLADENVITR-NELML	813
Db	455	VCDQP---VTCPTPKPDQVCGTDNQTY-----ASSCHLTATRCLEGTKKGHQLOL	503
QY	814	EETRNTITVPA-----SFMLRLMASLNHNRADRLEGDSRSGSGGENENE	857
Db	504	DYFGACKSIPTCTDPEVFIQFPLRWDRWLKNYLMOLEYEAN-SEHAGYLENEQ	553

## RESULT 19

US-08-320-559-28  
; Sequence 28, Application US/08320559  
; Patent No. 7,600,000

Patent No. 5633135  
GENERAL INFORMATION:

;  
; APPLICANT: Croce, Carlo  
; APPLICANT: Canaan, Eli

APPLICANT: Cahaani, Eli  
TITLE OF INVENTION: Diagnostic  
TITLE OF INVENTION: Diagnostic

TITLE OF INVENTION:	Detecting
TITLE OF INVENTION:	Results

TITLE OF INVENTION:	ALL-
NUMBER OF SEQUENCES:	44

SEQUENCES: 44  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock

ADDRESSEE: Woodcock Wa  
STREET: One Liberty Pl  
CITY: Pitts

CITY: Philadelphia  
STATE: PA

COUNTRY: USA  
ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: DISKETTE,  
COMPUTER: IBM PS/2

OPERATING SYSTEM: PC-DOS  
SOFTWARE: WORDPERFECT

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER:

us-09-781-693a-2.ra1

Wed Mar 26 09:38:53 2003

```

; FILING DATE: 435
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: US/08/062,443
; FILING DATE: 14 MAY 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/971,094
; FILING DATE: 30-OCT-92
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/888,830
; FILING DATE: 27-MAY-92
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/805,093
; FILING DATE: 11-DEC-91
; APPLICATION DATA:
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJU-0855
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1187 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-320-559-28

Query Match 3.7%; Score 165.5; DB 1; Length 1187;
Best Local Similarity 20.4%; Pred. No. 0.00029;
Matches 125; Conservative 80; Mismatches 218; Indels 189; Gaps 26;

QY 279 LFDPRDARELKTSAERREELRQPPVKRLRGDNDSTGP-----RARPESEREED 332
DB 240 LVAPAPPSQTPFPLSLPSKSVAMQKQYAVVRPM-DGQDQAPSPSELKPLPDYRQOT 298
QY 333 GEQS-----PNVSLQRMS-----DMLSRW----- 352
DB 299 FEKTLKVPKAKLTKLKMPQSVQETYNVHCVEELKEMTHSWPPLTAIHTPSTAE 358
QY 353 -----FEASVAVQSNRGRSRPRGGTQSQDSTLTLPVSPSSPLEVSEAMEVD 402
DB 359 PSKPPPTKDSQHVSVTQKQYDTSSKTHSNQOQTSSM-----LEDDQLSDSE--D 411
QY 403 TPAPQFLQ-PSTSTMSQAQHSSTSTESPHSTPLSSPDSEQRQVSEAGSHHTHQSDN 461
DB 412 SDSRQTEPKPPSSAPPSPQSLPEVASAHS-----SSAESSTSDSS-----SDS 460
QY 462 NNEKLSKPGTGPVLSIHYSTEGTTTSTIKLNFDTDW-----SSIASSSRGIGS----- 511
DB 461 ESESSSDSENEPLETAPPEPEPTTNKWLQ-----DNWLTKVQSPAAPPEGPRSTPPRR 517
QY 512 HCKSEGOEESVQ--SSVQPPEDSETKAPESSE-----DVTKYQGVSAENPVENHN 565
DB 518 HPESKSSDSATSQHSSEKDPKPKSSKAPAPPEAPHGPKRSCQKSPAQEPQOQ-- 575
QY 566 IQSDKFTAKPL--DSNNGERDNLNDRSCV-----PEESASS 602
DB 576 -TVGKQPKPKVKAARAGSRSTSLQGEREPGLLPYGRDQTSKDKPKVTKGRPRRAASN 634
QY 603 E-KAKEPETSQDTSTES---ATNENNTNPEPQFQ-----TEATGPSAHEE 643
DB 635 EPKPAVPPSPSEKKHKLSPAPSKALSGPEPAKDNVEDRTPHFALVPLTESQGP-PHSG 693
QY 644 TSTRDSALQDT-----DDSDDDPVLIP----- 665
DB 694 SGRSTGCCQAVVQEDSRKRLPLPLRDTKLPLSDRTPPPQSLMVKITDILLSRIQOP 753
QY 666 ---GARYRAG-----PGDRSNAVARIQEFRR-RKERKEMEELDTLNIRPLVKMYKG 715
DB 754 PGKGRQKAEKOPKPPAGKHKSEKSSSSSKLAKKRGAEKDCDNKKIRLEIKSQ 813

```

Query Match 3.7%; Score 165.5; DB 3; Length 1187;

```

QY 716 HRNRTMIKEAN 727
DB 814 SSSSSSHKSS 825

RESULT 20
US-08-545-860D-28
; Sequence 28, Application US/08545860D
; Patent No. 6040140
; GENERAL INFORMATION:
; APPLICANT: Croce, Carlo
; APPLICANT: Canaan, Eli
; TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1 Region
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
; ADDRESSEE: No. 6040140ris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/545,860D
; FILING DATE: 07-MAR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/04496
; FILING DATE: 22-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/10930
; FILING DATE: 09-DEC-1992
; APPLICATION NUMBER: US 08/327,392
; FILING DATE: 19-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/320,559
; FILING DATE: 11-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/062,443
; FILING DATE: 14-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/971,094
; FILING DATE: 30-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/888,839
; FILING DATE: 27-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/805,093
; FILING DATE: 11-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca Esq., Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJU-1262
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1187 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-545-860D-28

```



```

Best Local Similarity 20.4%; Pred. No. 0.00029;
Matches 125; Conservative 80; Mismatches 218; Indels 189; Gaps 26;

QY 279 LDPKDDTARELKTPSAEERREELRQPPVKRLRGDWSDTGP-----RARPESERERD 332
Db 240 LVAPAPPSQTFFPPSLPSKVAMQOKTAYVRPM-DGQDQAPSESPELKPLPDIYRQQT 298
QY 333 GEQS-----PNVSLMORMS-----DMLSRW-----SDS 352
Db 299 FEKTDLVKPAKAKUTKLKMPFSQVEQTSYNEVHCVEILKEMTHSWPPPLTAHTPSTAE 358
QY 353 -----FEASEVAQSNRGRSRPRGGTQSQSDISTLPTVPSPDLEVSETAMEVD 402
Db 359 PSKFPPTKDSQHVSVYTONQKQYDTSKTHSNSQOQTSSM-----LEDDQLQSDSE--D 411
QY 403 TPAEQFLQ-PSTSTMSAQAHSTSTSPHSTPLSSPDSEQRQSVASGHHTHQSDN 461
Db 412 SDSEQTPKPPSSAPPAPQSLPEPVAHSAH-----SSAESESTSDSDS-----SDS 460
QY 462 NNEKLSPKPGTGPVLSLHYSTEGTTSTIKLNFTEW-----SSIASSSRGIGS----- 511
Db 461 ESESSSDSEENEPLTPAPEPEPTTNKWL-----DNLTKVSOQAPAPPEGRSTPEPRR 517
QY 512 HCKSEGOESFVQ--SSVQPPGDSSETKAPESE-----DVKYQEGVSAENPVENHIN 565
Db 518 HPESKSSDSATSQHSESKDPPKSSKAPRAPPEAPHGKRSCOKSPAQOEPPQO-- 575
QY 566 ITQSDKFTAKPL--DSNNGERDNLDRSCGV-----PEESASS 602
Db 576 -TVGTPKPKPKVAKASARAGRTSLQGEREGLLPYGSRDQTSKDKPKVKTKGRPRAASN 634
QY 603 E-KAKEPETSQDTSTES---ATNENNTNPEPQFQ-----TEATGPSAHEE 643
Db 635 EPKPAVPPSEKKHKSLSLAPSKALSGPEAKDNVEDRTPHEFALVPLTESQGP-PHSG 693
QY 644 TSTRDSALQDT-----DSDDDDPVLIP----- 665
Db 694 SGRTSGROAVVQEDSRDKRLPLRLDRTKLLSPLRDTPPQSLMWKITLDLLSRIPQ 753
QY 666 ---GARYRAG-----PGDRRSARVARIQEFRR-RKERKEMEELDTLNIRRLPVKMYKG 715
Db 754 PGKSGRQKAEQKPPAGKHKSEKSSSSSLAKKRKGEARDCDNKKIRLEKEIKSQ 813
QY 716 HRNSRTMIKEAN 727
Db 814 SSSSSSSSHKSS 825

RESULT 21
PCT-US94-04496-28
; Sequence 28, Application PC/TUS9404496
; GENERAL INFORMATION:
; APPLICANT: Croce, Carlo
; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
; TITLE OF INVENTION: For Detection and Treatment of Acute Leukemias
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/04496

```

```

; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca Esq., Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJJU-1242
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1187 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US94-04496-28

```

```

Query Match 3.7%; Score 165.5; DB 5; Length 1187;
Best Local Similarity 20.4%; Pred. No. 0.00029;
Matches 125; Conservative 80; Mismatches 218; Indels 189; Gaps 26;

QY 279 LDPKDDTARELKTPSAEERREELRQPPVKRLRGDWSDTGP-----RARPESERERD 332
Db 240 LVAPAPPSQTFFPPSLPSKVAMQOKTAYVRPM-DGQDQAPSESPELKPLPDIYRQQT 298
QY 333 GEQS-----PNVSLMORMS-----DMLSRW-----SDS 352
Db 299 FEKTDLVKPAKAKUTKLKMPFSQVEQTSYNEVHCVEILKEMTHSWPPPLTAHTPSTAE 358
QY 353 -----FEASEVAQSNRGRSRPRGGTQSQSDISTLPTVPSPDLEVSETAMEVD 402
Db 359 PSKFPPTKDSQHVSVYTONQKQYDTSKTHSNSQOQTSSM-----LEDDQLQSDSE--D 411
QY 403 TPAEQFLQ-PSTSTMSAQAHSTSTSPHSTPLSSPDSEQRQSVASGHHTHQSDN 461
Db 412 SDSEQTPKPPSSAPPAPQSLPEPVAHSAH-----SSAESESTSDSDS-----SDS 460
QY 462 NNEKLSPKPGTGPVLSLHYSTEGTTSTIKLNFTEW-----SSIASSSRGIGS----- 511
Db 461 ESESSSDSEENEPLTPAPEPEPTTNKWL-----DNLTKVSOQAPAPPEGRSTPEPRR 517
QY 512 HCKSEGOESFVQ--SSVQPPGDSSETKAPESE-----DVKYQEGVSAENPVENHIN 565
Db 518 HPESKSSDSATSQHSESKDPPKSSKAPRAPPEAPHGKRSCOKSPAQOEPPQO-- 575
QY 566 ITQSDKFTAKPL--DSNNGERDNLDRSCGV-----PEESASS 602
Db 576 -TVGTPKPKPKVAKASARAGRTSLQGEREGLLPYGSRDQTSKDKPKVKTKGRPRAASN 634
QY 603 E-KAKEPETSQDTSTES---ATNENNTNPEPQFQ-----TEATGPSAHEE 643
Db 635 EPKPAVPPSEKKHKSLSLAPSKALSGPEAKDNVEDRTPHEFALVPLTESQGP-PHSG 693
QY 644 TSTRDSALQDT-----DSDDDDPVLIP----- 665
Db 694 SGRTSGROAVVQEDSRDKRLPLRLDRTKLLSPLRDTPPQSLMWKITLDLLSRIPQ 753
QY 666 ---GARYRAG-----PGDRRSARVARIQEFRR-RKERKEMEELDTLNIRRLPVKMYKG 715
Db 754 PGKSGRQKAEQKPPAGKHKSEKSSSSSLAKKRKGEARDCDNKKIRLEKEIKSQ 813
QY 716 HRNSRTMIKEAN 727
Db 814 SSSSSSSSHKSS 825

```

```

RESULT 22
US-08-320-559-26
; Sequence 26, Application US/08320559
; Patent No. 563135
; GENERAL INFORMATION:
; APPLICANT: Croce, Carlo
; APPLICANT: Canaani, Eli

```

Query Match	3.7%;	Score	165.5;	DB	1;	Length	1210;
Best Local Similarity	20.4%;	Pred. No.	0.0003;				
Matches	125;	Conservative	80;	Mismatches	218;	Indels	189;
Gaps							
26;							
QY	279	LFDPKDDTARELKTPTSAERREERELQPPVKRLRGDWSDTGP-----RARPESERERD	332				
Db	263	LVAPAAQPPSTFPPELSFKSVAAQQKPTAYVRPM-DQDQAQPSSEPKLPEDYRQQT	321				
QY	333	GEQS-----PNVSLMQRMS-----DMLSRW-----	352				
Db	322	FEKTDLKVPAAKATLTKMPQSVEQTYSNEVHCVEEILKEMTHSWPPPLTAIHTPSTAE	381				
QY	353	-----FEASEVAQSNKGRGRPRGGTQSODISTLPTVSSPDLEVSAMEYD	402				
Db	382	PSKEFPPTKDSOHVSSVTQKQYDTSKTHSNQOQTSSM-----LEDDQLQSDSE--D	434				
QY	403	TPAEQFLQ-PSVSSNTMSQAQAHSTGPTSPHSTPLSSPDSPQSVASGHHHTHQSDN	461				
Db	435	SDSEQTPEKPPSSAPPSAPQSLPEEVAHAHS-----SSAESESTSDSS-----SDS	483				
QY	462	NNEKLSKPKGTGTEPVLUSLAYTEGTTTSTIKLFTDEW-----SSIASSSRGIGS-----	511				
Db	484	ESESSSDSEENEPLETPAPEPPPTTNKWL---DNWLTKVSQAAPPEGGPSTPEPPR	540				
QY	512	HKSEQGEBSFVQQ---SSVQPPGEDSETKAPESSE-----DVTKYQGVSAENPVENHIN	565				

[illegible]



us-09-781-693a-2.ra1

Wed Mar 26 09:38:53 2003

112 VSCGDIYETVNEODATNROCQFTCHYGTTEIMTVNDPPTFLSCGDCGTVMW---168  
402 SENKDAKAVSLLNNCKDAEYKYCD--CKHTTTL--VKSVLN-----GNDNTIKEKE 450  
169 ---PDRITKISCTKE-----DKDDILINCRAATSAICPPPIPYLAVGCDSSV 216  
451 HIDLDDFSKFGCDKNVDNTNKVECKNPYIL-----STKDVCPVPRRQELCIGNID---502  
217 RIYDRML-----GTRATGN 231  
503 RIYDKNLMIKEHILATATAYESRILKRYKNKDKVECKIINKTFADIRDIIGTDYND 562  
232 YAGRTTGMV---ARFIPSHLNKSCRV-----TSICYSEDG 265  
563 LSNKLVGKINTNSKYV--HRNKKNDKLFDRDWMVKIKOVYVNSWVFKDKTVCKEDDI 620  
266 QELVYSYSDYIYLFDPKDDTARELKTSAERREELPQPVK-----RLRLGDM--SD 318  
621 ENI-----PQFRWFSEWDDYCYQDKTKMIETLVKVECKEPCEDDNCCKSCSYKEWISK 675  
319 TGPRAPESERERDEQSPNVSLMQMSDM-----LSRWFEASEVAQSNRGRSRRPG 373  
676 KKEEYNKQAKQOEYQKGNMYSEFKSIKPEVYLKRYSEKSNLFEDEFKEB-----730  
374 GTSQSDISTLTPVSPDLEVEVETAMEVDTPAQFLQSTSTMSQAQAHSTSP--TSP 431  
731 --LHSDYKNKCTM--CP--EVKDVPIIIRNEQISQEAPEENTELAHRTETPSISEGP 784  
432 HSTPLLSSPDSEQORQSVASGHHTHQSDNNNEKISPKPGTGEVLSLHYSTEGITSTI 491  
785 -----KGNEQKERDDDSLKISVSPENSRP-----ETDAKDTSL 819  
492 -----KLNEDWSSSTASSRSGIGSHCKSEGOEESFVPOSSVQ 529  
820 LKLKGDVDISMPKAVIGSSPNDNINTEQDNLIS-----GVNSKPLSD-----DVR 865  
530 PPEGSDSEKAPESSEEDVTYQEGVSAENPVENHINTQGDKFTAKPLDSNG-----ER 584  
866 PDKKELEDQNSDESEETVNH--ISKPSINNGDDSGSGSATVSESSSNTGLSDDDR 922  
585 NDMLDRSCGVPEESASSEKAKETSDOTSTES-----ATNENNTNPEQFQTEATG 637  
923 NGDTFVRT-----QDTANTVEDIRKENADKDEKGADEERHSTSESLSPEKMLTDNEG 978  
638 PSA--HEE--TSTRDSA-----LQDT-----DD-----656  
979 GNSLNHEEVKEHTSNDSNVQSGGIVNMVNEKLDKTLNPSSSLDEGKAHEELSEPNLS 1038  
657 SDDDPVLIPGARYRAGPGDRSAVARIOEFRR--RKERKEMEELDTLNIIRPLVKMYK 714  
1039 SDQDMSNTP-----GPLDNTS-----EETTERISNNEYKVNEREDERTLTKEYEDIVLK 1087  
715 GHNSR-----TMKEANFWANFVMSGDCGHIFWDRHTAHLMLLEAD 760  
1088 SHWNRESDDGELYDENSDLSLVNDESE--DAEAKMGNDTSEM---SHNSQH---IESD 1139  
761 -----NHVNCLOPHFPDPTLASSGIDYDI-----KIWSPLEESRIFN--798  
1140 QOKNDMKTYGDLGTHVQNEIS-----VPVTGEIDELKRESKESKI--HKADEERLSHTD 1192  
799 -RKLADDEVITRNELMLEETRN 818  
1193 IHKINPDRNSNTHLKDNRN 1213

Db 717 SGRSRTSGCRQAVVQEDSRKRLPLRDTKLLSLRDTPTPPQPSLMVKITLDDLSRIPQ 776  
QY 666 ---GARYRAG-----PGDRRSARVARIOEFRR--RKERKEMEELDTLNIIRPLVKMYK 715  
Db 777 PGKSGRQKAEKQPPAGKKSSEKSSSSSLAKKRKGEAERCDNKRKIRLEKEIKSQ 836  
QY 716 HRNSRTMIKEAN 727  
Db 837 SSSSSSHKES 848  
RESULT 25  
US-08-568-459A-4  
Sequence 4, Application US/08568459A  
Patent No. 5849306  
GENERAL INFORMATION:  
APPLICANT: Sim, Kim L.  
APPLICANT: Chitnis, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhaun  
APPLICANT: Wellens, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
AND PLASMODIUM FALCIPARUM ERITHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe Martens Olson & Bear  
STREET: 620 Newport Center Drive 16th Floor  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/568,459A  
FILING DATE: 07-DEC-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelsen, Ned  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NIH121.001CPI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1435 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Plasmodium falciparum  
US-08-568-459A-4  
Query Match 3.7%; Score 165.5; DB 2; Length 1435;  
Best Local Similarity 18.1%; Pred. No. 0.00039;  
Matches 188; Conservative 148; Mismatches 330; Indels 375; Gaps 53;  
QY 36 EFIOBKLEATLVHDGCVNTIC-----WNDTGEYILSGSDDTKLVISN 79  
Db 290 EFERKL--WEAMLSKNNINN--CKNIQBELQITQWKEWH--GEFLERDNRSKL-----341  
QY 80 PYSRKVLVTTIRSG-----HRANFSKAF-----LPCTNDK-----QI 111  
Db 342 PKSKCKNNTLYACEKECIDPCMRYDNIIRSFHWHTLSKEYETQKVPKNAENYLIKI 401

RESULT 26  
US-08-487-826B-4  
Sequence 4, Application US/08487826B  
Patent No. 5993827  
GENERAL INFORMATION:  
APPLICANT: Sim, Kim L.  
APPLICANT: Chitnis, Chetan  
APPLICANT: Miller, Louis H.

```

Query Match          3.7%; Score 165.5; DB 2; Length 1435;
Best Local Similarity 18.1%; Pred. No. 0.00039;
Matches 188; Conservative 148; Mismatches 330; Indels 375; Gaps 53;

QY 36 EFTQRLKLEATLNVHDCVNTIC-----WNDTGEYILSGSDDTKLVISN 79
Db 290 EREKL-WEAMLSEHKNNINN-CKNIPOBELQITOWIKEWH--GEFLLRDNRSKL--- 341

QY 80 PYSRKVLVTIRSG-----HRANIFSAKF-----LPCNDK-----QI 111
Db 342 PKSKCKNNNTLYEACEKECIDPCMKYRDWIIRSKFEWHTLSKETQKVFNENAYELIKI 401

QY 112 VSCSGDGVIFYNTVEQDANTRCOQFTCHGYCTTYETMTVPNDPYTFLSCGEDGTVRW--- 168
Db 402 SENKNDAKVSLLNCCDAEYSKYCD--CAHTTTL-VKSVLN-----GNONTIKEKRE 450

QY 169 ---FDRTRKTSCTKE-----DCKDILINCRRAATSAICPPYPLVAVGCSDDSV 216
Db 451 HIDLDDFKFGCDKNSVDTNTKVWECKNPYL-----STKDCVCPRQELCLGNID--- 502

QY 217 RIYDRML-----GTRATCN 231
Db 503 RIYDNLLMIKEHILAIAIYSRILKRYKKNDDKEVKCIINKTFADIRDIIGGTYWND 562

QY 232 YAGRTTGMV---ARFTPSHNNKSCRV-----TSLCYSEGD 265
Db 563 LSNRKLGVKINTNSKYV--HRNKNKDLFRDEWVKVKKDYNNVISWFKDTVCCKEDDI 620

QY 266 QEILVYSDDYIYLPDKDDTARELKPYSAPERREELQPPVK-----RLRLGDW-SD 318
Db 621 ENI-----POQFRWFSEWGDDYCDKTKMTETLTVCKECPCEDDNCKSCNYSKEWLK 675

```

585 NDNLDSGCVPEESASSEKAPETSDOTSTES-----ATNENTNPEPQFQTEATG 637  
 923 NGDTFVNT-----QDANTEDVIRKENADKDEDEKGADEERHSTSESLSPEKMLTDNEG 978  
 638 PSA--HEE--TSTRDSA-----LQDT-----DD-----656  
 979 GNSLNHEEVKEHTSNSDNVQSGGIVNMNVEKELKDTLENPSSSLDEGKAHELSPEPNS 1038  
 657 SDDPVLIPGARYRAGCDRRSAVARIQEFRR--RKERKEMEELDTLNIRRLPLVMVYK 714  
 1039 SQDMSNTP-----GPDNTS-----EETTERISNNEYKVNEREDERTLTKEYEDIVLK 1087  
 715 GHNSR-----TMKEANFWGANFVMSGDCGHIFIDRHTAHLMLLEAD 760  
 1088 SHMNSDDGELYDENSJLSTVNDESE--DAEAKMKND7SEM--SHNSQH---TESD 1139  
 761 -----NHVNCLOPHPPDPFILASSGIDYDI-----KIWSPLEESRIFN-- 798  
 1140 QOKNDMKTVDLGTTHVQNEIS-----VPTGEIDKLRRESKESKI-HKAEERLSHTD 1192  
 799 -RKLADDEVITRNLMEETRN 818  
 1193 IHKINPEDRNSNTLHLKDIRN 1213

RESULT 28  
 US-08-961-083-94  
 ; Sequence 94, Application US/08961083  
 ; Patent No. 6159469  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Choi et. al.  
 ; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
 ; NUMBER OF SEQUENCES: 452  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Human Genome Sciences, Inc.  
 ; STREET: 9410 Key West Avenue  
 ; CITY: Rockville  
 ; STATE: Maryland  
 ; COUNTRY: USA  
 ; ZIP: 20850  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
 ; COMPUTER: HP Vectra 486/33  
 ; OPERATING SYSTEM: MSDOS version 6.2  
 ; SOFTWARE: ASCII Text  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/961,083  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Brookes, A. Anders  
 ; REGISTRATION NUMBER: 36,373  
 ; REFERENCE/DOCKET NUMBER: PB340P2  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (301) 309-8504  
 ; TELEFAX: (301) 309-8512  
 ; INFORMATION FOR SEQ ID NO: 94:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 278 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-961-083-94

Query Match 3.6%; Score 164; DB 4; Length 278;  
 Best Local Similarity 22.2%; Pred. No. 4.2e-05;  
 Matches 66; Conservative 50; Mismatches 117; Indels 64; Gaps 10;  
 387 PSSPDLEVSETA-----MEVDYTAELQFLOPSTSTMSQAHSSTSPTE 429

3.7%; Score 165.5; DB 4; Length 1435;  
 Best Local Similarity 18.1%; Pred. No. 0.00039;  
 Matches 188; Conservative 148; Mismatches 330; Indels 375; Gaps 53;  
 36 EFIOQLKLEATLNVDHGVNTIC-----WNTGEXYLSGSDDTKLVIN 79  
 290 EFREKL-WEAMLSBKNNINN-CNKIPQEEELQITQWIKWH--GEFLLERDNRSKL--- 341  
 80 PYSRKVLTRISG-----HRNIFSAKF-----LPCNDK---QI 111  
 342 PKSCKNNTLYEACEKICDPCMYRDLWILRSKEFWHTLSKEYETQKVPKRENAENYLIK 401  
 112 VSCSGDGVIFYTNYVODAEATNRQCOFTCHGYTTEIMTVNPDYTFILSCGEGTVRW--- 168  
 402 SENKNDARVSLNLCNDAEYKCYCD--CKHTTTL-VKSVLN-----GNDNTIKERE 450  
 169 ---FTRIKTSCTKE-----DCKDDILNCRRATSVACPPPIPYILAVGCSDDSV 216  
 451 HDLDDFSKFGCDKNSVDTNTKWECKNPYL---STKDCVCPRRQELCLGNID--- 502  
 217 RIYDRML-----CNRATGN 231  
 503 RIYDKNLLMIKEHILAIAIESRLKRYKKNDDKEVCCKIINKTFADIRIDIGTDVND 562  
 232 YAGRTGTGMV---ARFIPSHLNKSCRV-----TSLCYSDG 265  
 563 LSNRLVKGINTNSKIV--HRNKNKDLFDEWVKVKKOVNWNVISWVFKDKTVCKEDDI 620  
 266 QETLVSYSDYIYLFDPKDDTARELKTSAERREELQPPVK-----RLRLGDM--SD 318  
 621 ENI-----PQFFRWFSWGGDYCDQKTKMIETLKVECKEPCEDDNCCKSKYKEWISK 675  
 319 TGPRAPRESEERERDQSNVSLMQMSDM-----LSRWFEASEVAOSNRGRGRSPRG 373  
 676 KKEYNKQAKQYQYQKGNMYKMYSEFKIKPEVYLKYSKCSNLNFEDEFKEE----- 730  
 374 GTSQSDISTLPTVPSPPDLFVSTAMEVDTAPQFLOPSTSTMSQAHSSTSP--TESP 431  
 731 --LHSDYKNNCTM--CP--EVKDVPISTIRNNEQTSQEAPEENTEIAHRTETPISSEGP 784  
 432 HSTPLSSPDSEQRQVSQVSEAGHHTHQSDNNNEKLPKPGTGPVLSLHYSYEGTGTSTI 491  
 785 -----KGNQKERRDDLSLSKISVSPNSRP-----ETDAKDTSNL 819  
 492 -----KLNFTDESSITASSRGGISHCKSEGBQESVPOSSVQ 529  
 820 LKLKGDYDIMPXKAVIGSSPNDNINVTQEDNIS---GVNSKPLSD-----DVR 865  
 530 PPGDSKAPESSESDYTKYQCVSAENPVENHINITQSDKFTAKPLDSNSG-----ER 584  
 866 PDKLEEDQNSDESEETVNH---ISKSPSINNGDDSGSGSAIVSESSSNTGLSDDDR 922

Query Match 3.7%; Score 165.5; DB 4; Length 1435;  
 Best Local Similarity 18.1%; Pred. No. 0.00039;  
 Matches 188; Conservative 148; Mismatches 330; Indels 375; Gaps 53;  
 36 EFIOQLKLEATLNVDHGVNTIC-----WNTGEXYLSGSDDTKLVIN 79  
 290 EFREKL-WEAMLSBKNNINN-CNKIPQEEELQITQWIKWH--GEFLLERDNRSKL--- 341  
 80 PYSRKVLTRISG-----HRNIFSAKF-----LPCNDK---QI 111  
 342 PKSCKNNTLYEACEKICDPCMYRDLWILRSKEFWHTLSKEYETQKVPKRENAENYLIK 401  
 112 VSCSGDGVIFYTNYVODAEATNRQCOFTCHGYTTEIMTVNPDYTFILSCGEGTVRW--- 168  
 402 SENKNDARVSLNLCNDAEYKCYCD--CKHTTTL-VKSVLN-----GNDNTIKERE 450  
 169 ---FTRIKTSCTKE-----DCKDDILNCRRATSVACPPPIPYILAVGCSDDSV 216  
 451 HDLDDFSKFGCDKNSVDTNTKWECKNPYL---STKDCVCPRRQELCLGNID--- 502  
 217 RIYDRML-----CNRATGN 231  
 503 RIYDKNLLMIKEHILAIAIESRLKRYKKNDDKEVCCKIINKTFADIRIDIGTDVND 562  
 232 YAGRTGTGMV---ARFIPSHLNKSCRV-----TSLCYSDG 265  
 563 LSNRLVKGINTNSKIV--HRNKNKDLFDEWVKVKKOVNWNVISWVFKDKTVCKEDDI 620  
 266 QETLVSYSDYIYLFDPKDDTARELKTSAERREELQPPVK-----RLRLGDM--SD 318  
 621 ENI-----PQFFRWFSWGGDYCDQKTKMIETLKVECKEPCEDDNCCKSKYKEWISK 675  
 319 TGPRAPRESEERERDQSNVSLMQMSDM-----LSRWFEASEVAOSNRGRGRSPRG 373  
 676 KKEYNKQAKQYQYQKGNMYKMYSEFKIKPEVYLKYSKCSNLNFEDEFKEE----- 730  
 374 GTSQSDISTLPTVPSPPDLFVSTAMEVDTAPQFLOPSTSTMSQAHSSTSP--TESP 431  
 731 --LHSDYKNNCTM--CP--EVKDVPISTIRNNEQTSQEAPEENTEIAHRTETPISSEGP 784  
 432 HSTPLSSPDSEQRQVSQVSEAGHHTHQSDNNNEKLPKPGTGPVLSLHYSYEGTGTSTI 491  
 785 -----KGNQKERRDDLSLSKISVSPNSRP-----ETDAKDTSNL 819  
 492 -----KLNFTDESSITASSRGGISHCKSEGBQESVPOSSVQ 529  
 820 LKLKGDYDIMPXKAVIGSSPNDNINVTQEDNIS---GVNSKPLSD-----DVR 865  
 530 PPGDSKAPESSESDYTKYQCVSAENPVENHINITQSDKFTAKPLDSNSG-----ER 584  
 866 PDKLEEDQNSDESEETVNH---ISKSPSINNGDDSGSGSAIVSESSSNTGLSDDDR 922

```

Db 9 PESPDTVWSDKGEPEQVAPLPEYKGNIEQVKEPTVEKTEQGPXTEEV-----PVK 61
QY 430 SPHSTPLSSPDSEQRQSVASGHHTH--HQSDNNNEKLSPKPGTGPVLSLHYSTEGTT 487
Db 62 PTEETPVNNEGTEGTSIQEAENPVQPAESTTTSEKRVSP-----DTSSKN 108
QY 488 TSTIKLNTDEWSSSTASSRGIGSHCKSGEQEESFVQSSVQPPGDSSTKAPESSEDEV 547
Db 109 TEVSSNPSDSTSTGESKNPKHNDKNSSEKT-VEEVVNPNEGTVGTSNQETEKPV 167
QY 548 TKYQGVSAENPVENHINITQSDKFTAKPLDSNSGERDNLNDRSCGVPPEASSEK--- 604
Db 168 OPAEE-----TQFN--SGKIANENTGEVSKPKPSDSKPPV-EESNQPEKNGT 210
QY 605 AKPEPSTQSTESATNNTNPEQFQTEATGFSAAHETSTROSALQDQDSDDDP 661
Db 211 ATKPNSGNTTSEN---GQTEPEPSNGNSTEDVSTESNTSNGNEIKQENELDP 263

RESULT 29
US-08-190-802A-63
; Sequence 63, Application US/08190802A
; Patent No. 5519003
; GENERAL INFORMATION:
; APPLICANT: Mochly-Rosen, Daria
; APPLICANT: Ron, Dorit
; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: P.O. Box 60850
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-0850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/190.802A
; FILING DATE: 01-FEB-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 8600-0139
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 713 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: TUP1, Fig. 46
;
Query Match 3.4%; Score 154; DB 1; Length 713;
Best Local Similarity 20.2%; Pred. No. 0.0011;
Matches 93; Conservative 77; Mismatches 186; Indels 104; Gaps 18;

```

```

QY 369 SRPRGGTSQSDISTLPVPSPPDEVSATAMEYDTPAQFLQSTSTMSAQAHSTSPST 428
Db 159 SRPNLVGSQLPTTLPLVPSNAQQQLPQQQLQOQQOQQPPQVSVAPLSNTAINGSPT 218

```

```

QY 429 ESPHST-PLSSPDSEQRQSVASGHHTH--HQSDNNNEKLSPKPGTGPVLSLHYSTEGTT 487
Db 219 SKETTTLPSVKAPESILKET-----EPENNNTSKIN---DTGSA-----TATTAT 260
QY 488 TSTIKLNTDEWSSSTASSRGIGSHCKSGEQEESFV-----QSSVOPP---EGDSET 537
Db 261 TATEIKPEEDATPASLH-----QDHYLPVYNQRANHSKPIPPFLDLDSDQS 309
QY 538 ---KAPEESSEDEVTKYQGVSAENPV-----NHINITQSDKFTAKPLDSNSGERDNLN 589
Db 310 VPDALKKQINDYIILYNPALPREIDVELHKSLDHTSVVCCVKF-----SNDGEYLATGC 363
QY 590 DRSCGVPPEASSEKAKEPETSQSTESATNNTNPEPQFQTEATGFSAAHETSTROS 649
Db 364 NKTTQVRYVSDGSLVARLSDSAAANNHRNSITENTTTSTDNNTWTTTTTTTTTTTAMTS 423
QY 650 ALQ-----DTDDSDDDPVLI-----PGARYRA-GPGDRRSAAVARIOEFRRRK 694
Db 424 AAELAKDVENLTSSSPSDLYIRVCFSPDGKFLATGAEDR---LIRIWDIENRK--- 476
QY 695 EMBELDTLNIIRPLVKMYKGHNSRTMIKEANFWGANF-----VMSGSDCGHIFWDR 748
Db 477 -----IVMILQGH-----EQDIYSLDYFFSGDKLVSGSDRTVRIWDL 514
QY 749 HTAELMLLEADNHVNVNCLQHPHPDPILASSGIDYDIKIW 788
Db 515 RTGQSLTSLIEDGVTTVAVSPGDGKYIAAGSLDRAVRVW 554

RESULT 30
US-08-477-346-63
; Sequence 63, Application US/08477346
; Patent No. 6262023
; GENERAL INFORMATION:
; APPLICANT: Mochly-Rosen, Daria
; APPLICANT: Ron, Dorit
; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morrison & Foerster
; STREET: 2000 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,346
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,072
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 2550-0025.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 713 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO

```

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500

```

;
; SOFTWARE: PATENTLII Release #1.07 (October 1997)
;
; CURRENT APPLICATION DATA:
;

```

; CURRENT APPLICATION DATA:

; ;  
CURR  
Soc



```

RESULT 34
PCT-US93-03077-1
; Sequence 1, Application PC/TUS9303077
; GENERAL INFORMATION:
; . APPLICANT: Board of Regents, The University of Texas System
; . APPLICANT: Gaynor, Richard B.
; . APPLICANT: Wu, Foon Kin
; TITLE OF INVENTION: PROTEIN CELLULAR FACTOR USEFUL FOR
; TITLE OF INVENTION: REGULATING GENE EXPRESSION
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; Zip: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/03077
; FILING DATE: 19930331
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

```

us-09-781-693a-2.rai

Wed Mar 26 09:38:53 2003

```

; APPLICATION NUMBER: US/07/862.025
; FILING DATE: April 2, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Kammerer, Patricia A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: UTDFD270PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-787-1540
; TELEFAX: 713-749-2679
;
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1093 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; PCT-US93-03077-1
;
; Query Match 3.4%; Score 154; DB 5; Length 1093;
; Best Local Similarity 20.3%; Pred. No. 0.0021;
; Matches 100; Conservative 81; Mismatches 186; Indels 126; Gaps 20;
;
; QY 269 LVSYSDYIYLFDPKDTARELTPTSAERRELQPPVKRLRLGWDSDTGRARPESE 328
; Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
; QY 89 VVDESEFFSAFLSPDTQVITQKSP-----VVSRPPAK-----SQRPPEE 128
; Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
; QY 329 RERDGEQSPNVSLMORMSDMLSRWFSEASVAQSNRGRSRPRGG-----TSQSDIS 381
; Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
; QY 129 VKSLSHSLHGQRTPTETTESQVKDSLCLVSGETLAAGTSSPKTEGKHEETVKNESDMK 188
; Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
; QY 382 TLPTVSPDLEVTAMVDTPAEQLQFQSTSTMSAQAHSTSSPESHPLSSPD 441
; Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
; QY 189 -VPTV---SLKVSSEVIDVKTMTESINTSQSLTATETKDIALEPKQKH----- 234
; Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
; QY 442 SEQQSVASGHHTHQSDNNKSLPKPGTGEVLSLHYSTEGTTTSTIKLNTFDWSS 501
; Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
; QY 235 -EDRUS-----NTPSPPVST-FSSGTSTSDIEV--LDHESV 267
; Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
; QY 502 IASSRGIGSHCKSGQESFVQSSVQPPGCD---SETKAPRESS-EDVTKYQEGVSA 556
; Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
; QY 268 ISESS-----ASSRQETDSSKSLHMQTSFQLLSASACPEVRLDDFKLTSCCS 319
; Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
; QY 557 ENPVENHINITQDFTAKPLDSSNGERNDLNRSCG-----VP-EESASSEKAKEPE 609
; Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
; QY 320 SDAFE-----RIDFSVQSLDSRS--VSEINSDDELSCGKGVLPVIVNSSTPKSTVE 371
; Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
; QY 610 TSDQSTESATNNNTNPPQFQTEATGPSA-----HEETSTRDSALQ 652
; Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
; QY 372 SAEGKSEE--VNETLVIPTTEAEEMESGRSATPVNCEQPDILVSTPNEGQTVLDKYAE 429
; Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
; QY 653 DTDDSDDDVLPFGARYRAGPGRDRSAVARIQFFRRRKERKEMELDTLNRPLVMV 712
; Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
; QY 430 QCEPAESQPEAL-----SEKEDVCKTVEFLNEKLEKREAQLL-SLSKEKALLEEA 478
; Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
; QY 713 YKGRHNSRTMIKE 725
; Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
; QY 479 FDLNKOEMFRVKE 491
; Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
;
; RESULT 35
; US-08-190-802A-202
; Sequence 202, Application US/08190802A
; Patent No. 5519003
; GENERAL INFORMATION:
; APPLICANT: Mochly-Rosen, Daria
; APPLICANT: Ron, Dorit
; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: P.O. Box 60850

```

```

; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-0850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/190,802A
; FILING DATE: 01-FEB-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 8600-0139
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 202:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: MUS MUSCULUS PROTEIN rII, Fig. 37
; US-08-190-802A-202
;
; Query Match 3.4%; Score 153.5; DB 1; Length 50;
; Best Local Similarity 50.0%; Pred. No. 2.1e-05;
; Matches 25; Conservative 13; Mismatches 11; Indels 1; Gaps 1;
;
; QY 741 GHFIWDRITAEHLMLLEAD-NHVVNCLQHPHFPDPFPLASSGIDYDIKIWS 789
; Db 1 GHFIWEKSCQIVQFLEADGGTINCIDSHYPVLPVLIASSGLDHEVKIWS 50
;
; RESULT 36
; US-08-477-346-202
; Sequence 202, Application US/08477346
; Patent No. 6262023
; GENERAL INFORMATION:
; APPLICANT: Mochly-Rosen, Daria
; APPLICANT: Ron, Dorit
; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morrison & Foerster
; STREET: 2000 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,346
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,072
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959

```

REFERENCE/DOCKET NUMBER: 2550-0025.20  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202) 887-1500  
 TELEFAX: (202) 887-0763  
 INFORMATION FOR SEQ ID NO: 202:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 50 amino acids  
 TYPE: amino acid  
 TOPOLOGY: unknown  
 MOLECULE TYPE: peptide  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 INDIVIDUAL ISOLATE: MUS MUSCULUS PROTEIN rII, Fig. 37  
 US-08-477-346-202

Query Match 3.4%; Score 153.5; DB 4; Length 50;  
 Best Local Similarity 50.0%; Pred. No. 2.1e-05;  
 Matches 25; Conservative 13; Mismatches 11; Indels 1; Gaps 1;

QY 741 GHIFWDRHTAHLMLLEAD-NHVVNCLQHPHPDPILASSGIDYDIKIWS 789  
 |||||: : : |||| :||: ||: |||||:|:||||  
 Db 1 GHIFWEKSCQIVQFLEADGGTINCIDSHPLPVLASSGLDHEVKIWS 50

## RESULT 37

US-08-473-089-202  
 Sequence 202, Application US/08473089  
 Patent No. 6342368

GENERAL INFORMATION:  
 APPLICANT: Mochly-Rosen, Daria  
 APPLICANT: Ron, Dorit  
 TITLE OF INVENTION: WD-40 - Derived Peptides and Uses  
 TITLE OF INVENTION: Thereof  
 NUMBER OF SEQUENCES: 265  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Morrison & Foerster  
 STREET: 2000 Pennsylvania Avenue, NW  
 CITY: Washington  
 STATE: DC  
 COUNTRY: USA  
 ZIP: 20006-1812

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/473,089  
 FILING DATE: 07-JUN-1995  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: MURASHIGE, KATE H.  
 REGISTRATION NUMBER: 29,959  
 REFERENCE/DOCKET NUMBER: 2550-0025.22

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202) 887-1500  
 TELEFAX: (202) 887-0763  
 INFORMATION FOR SEQ ID NO: 202:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 50 amino acids  
 TYPE: amino acid  
 TOPOLOGY: unknown  
 MOLECULE TYPE: peptide  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 INDIVIDUAL ISOLATE: MUS MUSCULUS PROTEIN rII, Fig. 37  
 US-08-473-089-202

Query Match 3.4%; Score 153.5; DB 4; Length 50;  
 Best Local Similarity 50.0%; Pred. No. 2.1e-05;  
 Matches 25; Conservative 13; Mismatches 11; Indels 1; Gaps 1;

QY 741 GHIFWDRHTAHLMLLEAD-NHVVNCLQHPHPDPILASSGIDYDIKIWS 789  
 |||||: : : |||| :||: ||: |||||:|:||||  
 Db 1 GHIFWEKSCQIVQFLEADGGTINCIDSHPLPVLASSGLDHEVKIWS 50

## RESULT 38

US-08-487-072A-202  
 Sequence 202, Application US/08487072A  
 Patent No. 6423684

GENERAL INFORMATION:  
 APPLICANT: Mochly-Rosen, Daria  
 APPLICANT: Ron, Dorit  
 TITLE OF INVENTION: WD-40 - Derived Peptides and Uses  
 TITLE OF INVENTION: Thereof  
 NUMBER OF SEQUENCES: 265  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Morrison & Foerster  
 STREET: 2000 Pennsylvania Avenue, NW  
 CITY: Washington  
 STATE: DC  
 COUNTRY: USA  
 ZIP: 20006-1812

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/487,072A  
 FILING DATE: 07-JUN-1995  
 CLASSIFICATION: 514  
 ATTORNEY/AGENT INFORMATION:  
 NAME: MURASHIGE, KATE H.  
 REGISTRATION NUMBER: 29,959  
 REFERENCE/DOCKET NUMBER: 2550-0025.20

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202) 887-1500  
 TELEFAX: (202) 887-0763  
 INFORMATION FOR SEQ ID NO: 202:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 50 amino acids  
 TYPE: amino acid  
 TOPOLOGY: unknown  
 MOLECULE TYPE: peptide  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 INDIVIDUAL ISOLATE: MUS MUSCULUS PROTEIN rII, Fig. 37  
 US-08-487-072A-202

Query Match 3.4%; Score 153.5; DB 4; Length 50;  
 Best Local Similarity 50.0%; Pred. No. 2.1e-05;  
 Matches 25; Conservative 13; Mismatches 11; Indels 1; Gaps 1;

QY 741 GHIFWDRHTAHLMLLEAD-NHVVNCLQHPHPDPILASSGIDYDIKIWS 789  
 |||||: : : |||| :||: ||: |||||:|:||||  
 Db 1 GHIFWEKSCQIVQFLEADGGTINCIDSHPLPVLASSGLDHEVKIWS 50

## RESULT 39

US-08-769-309A-5  
 Sequence 5, Application US/08769309A  
 Patent No. 5741890

GENERAL INFORMATION:  
 APPLICANT: Scott, John D.,  
 APPLICANT: Nauert, Brian J.,  
 APPLICANT: Klauck, Theresa M.  
 TITLE OF INVENTION: Protein Binding Domains of Gravin  
 NUMBER OF SEQUENCES: 24  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
 STREET: 6300 Sears Tower/233 South Wacker Drive

QY 771 P-----FDPILASGIDYDIKIW-----SPLSESRIFNRKLA 802  
 Db 873 OKAATEVSKELSESQVHMMAAAVADGTRAATIIERSPSWISASVTEPLEQVEAEAAULT 932  
 QY 803 DEVITRNLMELETRNTITVP 823  
 Db 933 EEVLER-EVIAEEEPPTVTEP 952

RESULT 40  
 US-08-994-570-5  
 : Sequence 5, Application US/08994570  
 : Patent No. 6090929  
 : GENERAL INFORMATION:  
 : APPLICANT: Scott, John D.,  
 : APPLICANT: Nauert, Brian J.,  
 : APPLICANT: Klauck, Theresa M.  
 : TITLE OF INVENTION: Protein Binding Domains of Gravin  
 : NUMBER OF SEQUENCES: 24  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
 : STREET: 6300 Sears Tower/233 South Wacker Drive  
 : CITY: Chicago  
 : STATE: Illinois  
 : COUNTRY: United States of America  
 : ZIP: 60606-6402  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: Patent In Release #1.0, Version #1.30  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/994.570  
 : FILING DATE:  
 : CLASSIFICATION:  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: No. 6090929and, Greta E.  
 : REGISTRATION NUMBER: 35,302  
 : REFERENCE/DOCKET NUMBER: 27866/33451  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: 312-474-6300  
 : TELEFAX: 312-474-0448  
 : TELEX: 25-3856  
 : INFORMATION FOR SEQ ID NO: 5:  
 : SEQUENCE CHARACTERISTICS:  
 : LENGTH: 1780 amino acids  
 : TYPE: amino acid  
 : TOPOLOGY: linear  
 : MOLECULE TYPE: protein  
 : US-08-769-309A-5

Query Match 3.4%; Score 152.5; DB 1; Length 1780;  
 Best Local Similarity 20.7%; Pred. No. 0.0058;  
 Matches 141; Conservative 249; Indels 205; Gaps 33;

QY 281 DPKDTARELTPS-----AERREELRPPVKRLRGDMSDTPRPPESEREDGEQS 336  
 Db 339 DGKAEVASEKLTAHQAPQPAESAHEP-----RLSAEY-----KVLPSEEQVSSQG 389  
 QY 337 PNVSLMQRMSDMLSNWFEEASVQSNRGRSRPRGCTSDISTLTPTVSPDLEVSE 396  
 Db 390 PS-----EERPAPLATEVDFDKIEVHOEE-----VVAEVHVSTVEERTEEQKTEVEE 436  
 QY 397 TAMEVDTPAEQFL-----QPSTSS-----TMSAQAHSTSSPTSPHSTP---LLSSPDSE 443  
 Db 437 TAGSV--PAELVGMDAEPQAPAKELVKETCVSGEDPTQGGADLSDEKVLSPKPPG 494  
 QY 444 QROSVASGHHTHQSDNNNEKLSPPGCTGEPVLSLHYSTEGTTSTIKLNFDEWSSIA 503  
 Db 495 VVSEVEML-----SSQERMKVQ---GSLPKLKF-----TSTGLKLSGKKQ-----532  
 QY 504 SSSRGIGSHCKSGQESFVQSSVOPPE---GDSETKAPESSE-----545

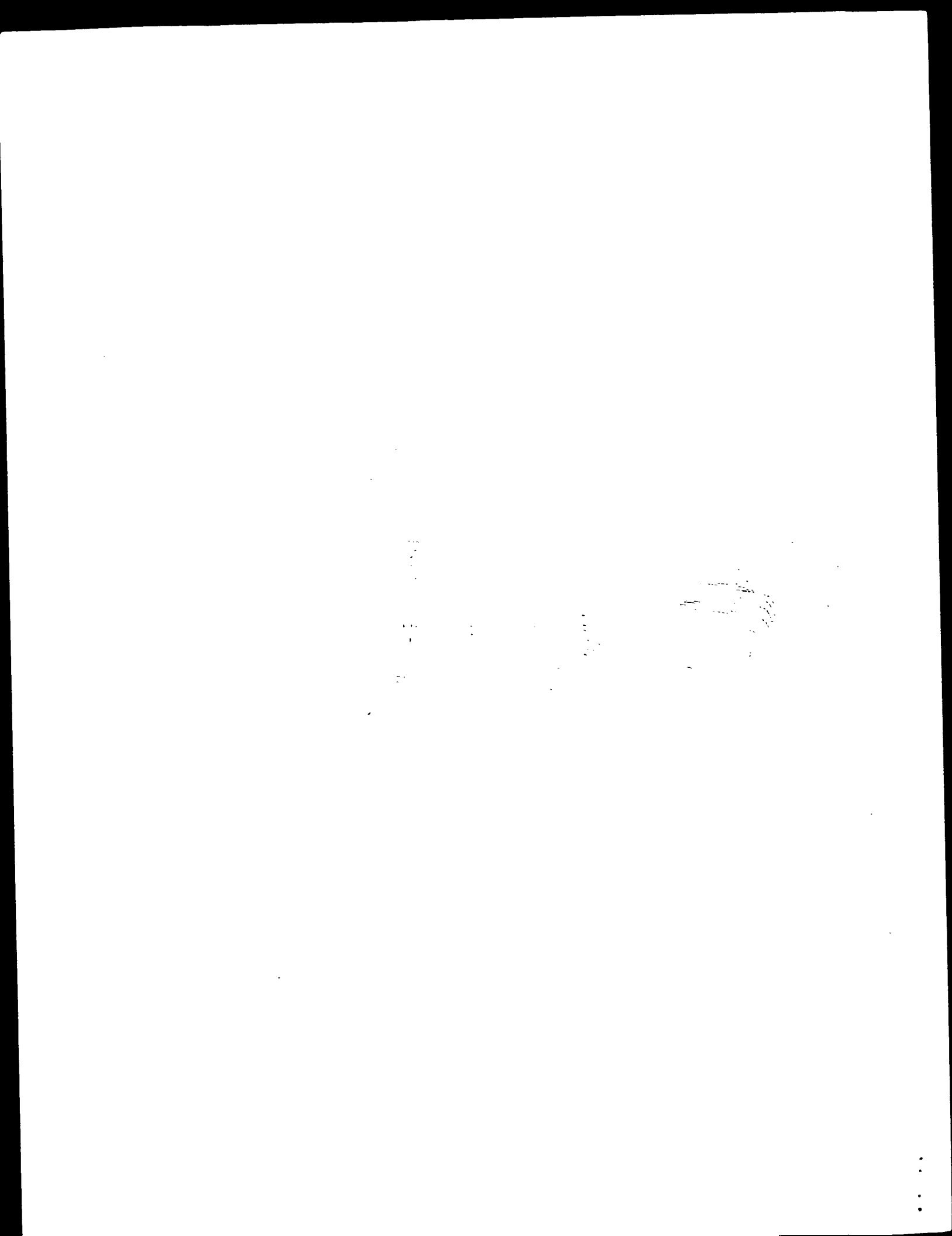
CITY: Chicago  
 STATE: Illinois  
 COUNTRY: United States of America  
 ZIP: 60606-6402  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/769.309A  
 FILING DATE:  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: No. 5741890and, Greta E.  
 REGISTRATION NUMBER: 35,302  
 REFERENCE/DOCKET NUMBER: 27866/33451  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 312-474-6300  
 TELEFAX: 312-474-0448  
 TELEX: 25-3856  
 INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1780 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-769-309A-5

Query Match 3.4%; Score 152.5; DB 1; Length 1780;  
 Best Local Similarity 20.7%; Pred. No. 0.0058;  
 Matches 141; Conservative 249; Indels 205; Gaps 33;

QY 281 DPKDTARELTPS-----AERREELRPPVKRLRGDMSDTPRPPESEREDGEQS 336  
 Db 339 DGKAEVASEKLTAHQAPQPAESAHEP-----RLSAEY-----KVLPSEEQVSSQG 389  
 QY 337 PNVSLMQRMSDMLSNWFEEASVQSNRGRSRPRGCTSDISTLTPTVSPDLEVSE 396  
 Db 390 PS-----EERPAPLATEVDFDKIEVHOEE-----VVAEVHVSTVEERTEEQKTEVEE 436  
 QY 397 TAMEVDTPAEQFL-----QPSTSS-----TMSAQAHSTSSPTSPHSTP---LLSSPDSE 443  
 Db 437 TAGSV--PAELVGMDAEPQAPAKELVKETCVSGEDPTQGGADLSDEKVLSPKPPG 494  
 QY 444 QROSVASGHHTHQSDNNNEKLSPPGCTGEPVLSLHYSTEGTTSTIKLNFDEWSSIA 503  
 Db 495 VVSEVEML-----SSQERMKVQ---GSLPKLKF-----TSTGLKLSGKKQ-----532  
 QY 504 SSSRGIGSHCKSGQESFVQSSVOPPE---GDSETKAPESSE-----545  
 Db 533 KGRGGD---EESGEHTQVPADSPDSQBEQKGESSASSPEEPTCLEKGLAEVQDQ 589  
 QY 546 -----DVTKYQGVSAENPVENHINTVQSDKFTAKPLDLSNGSRNDLNRSCGVPE 597  
 Db 590 EAEAGTADGKKKREGVT---PWASFKMVTPKKVRPP--SESDKEDLDKVKSATLSS 644  
 QY 598 -ESASSE-----KAKEPETSQDST-----ESATNENNTNPEPQFQ 632  
 Db 645 TESTASEMQEBMGVSEEPKPEPKRVDTSVSWEALICVSGSKKRRARRSSD-----698  
 QY 633 TEATGPSA---HEETSTRSALQDQDSD-----DPVLIPGARYRAGPGRRSA 679  
 Db 699 -EEGPKAMGDHQA---DEAGDKETGTDGILAGSQEHDPGGSSSPQAGSPTEGEG 754  
 QY 680 VARIOEPPR---RRKERKEMEELDUNI-----RRP-----LVKMVYKGRH 717  
 Db 755 VSTWSEFRKLVTPRKKSKELEKSEDSIAGSGVHEHSTPDTPEKESWSVSKFIQRR 814  
 QY 718 NSR-----TWIKANFWANFYMSGSDGHIIFWDRHTAEHLMLLEADNHNVCNLOPH 770  
 Db 815 KKRPGKQEQAPVEDAGPTGAN---EDDSQVAVVPLSEYDAVEREKMEAOQAQKGAQPE 872

Db 533 GKRCGGD---EEGSEHTQVPADSPDSOEQKQESSASSPEEPEITCLEKGLAEVQODG 589  
QY 546 -----DVTKYQGVSAENPVENHINITQDKFTAKPLDSNGERNDLNLDRSCGVPE 597  
Db 590 EAEGATSDGKKEGVT---PWASFKKMTVPKKVRVP--SESDKEDELKVKSATLSS 644  
QY 598 -ESASSE-----KAKPETSDQST-----ESATNENNTNPEPOFQ 632  
Db 645 TESTASEMQEEMKGVPEPKPEPKRVDTSVWEALICVGGSSKKRARRSSSD----- 698  
QY 633 TEATGPSA---HEETSTRDSALQDTSDD-----DPVLPGARYRAGPGDRRSA 679  
Db 699 -EGGPKAMGGDHOKA---DEACKKETGTDGILLAGSQEHDPGOGSSSPQAGSPTGEG 754  
QY 680 VARIOEFFR---RRKERKEMEELDTLNI-----RRP-----LVKMVYKGR 717  
Db 755 VSTWSEFKRLVTPRKKSLSKLEKSEDSIAGSGVEHSTPDTEPKKEBSWYSIKFIPGR 814  
QY 718 NSR-----TMIKEANFWGANFVMSGDCGHIFWDRHTAEHLMLLEADNHVYVNCLOPH 770  
Db 815 KRPDQKQEQAPVEDAGPTGAN--EDDSVPVAVVPLSEYDAVEREKMEAOQAKGAEQPE 872  
QY 771 P-----FDPILASSGIDYDIKIW-----SPLEESRIFNRKLA 802  
Db 873 OKAAATEVSKELSESQVHMAAAVADGTRAATILEERSPSWISASVTEPLEQVEAEALLT 932  
QY 803 DEVITRNLMLLEETRTITVP 823  
Db 933 EEVLER-EVIAEEEPPTVTEP 952

Search completed: March 18, 2003, 15:36:37  
Job time : 59 secs



GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 18, 2003, 15:29:40 ; Search time 15 seconds  
(without alignments)  
2377.976 Million cell updates/sec

Title: US-09-781-693a-2

Perfect score: 4527

Sequence: 1 MSRGSSPHLLWDVRRSLG.....RLGGDRSEGSQENENEDDE 860

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: LASTING - first 126 summaries

Database: SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	221.5	4.9	489	1	DMPL1_RAT
2	221	4.9	503	1	DMPL1_MOUSE
3	198.5	4.4	406	1	SR40_YEAST
4	190.5	4.2	1367	1	ANYH_YEAST
5	189	4.2	1142	1	MGCL1_HUMAN
6	188	4.2	1702	1	IGA2_HAEM
7	186	4.2	1694	1	IGA0_HAEM
8	184	4.1	1025	1	HIRA_FUGRU
9	177.5	3.9	1658	1	YM67_YEAST
10	175.5	3.9	1849	1	IGA4_HAEM
11	175	3.9	934	1	DSPP_MOUSE
12	174.5	3.9	787	1	NASP_MOUSE
13	174.5	3.9	2004	1	MOZ_HUMAN
14	173	3.8	2492	1	ATRX_HUMAN
15	172	3.8	513	1	DMPL1_HUMAN
16	171.5	3.8	1337	1	DXTX_MOUSE
17	171.5	3.8	1189	1	YJH6_YEAST
18	170.5	3.8	1092	1	NCA2_XENLA
19	169.5	3.7	1253	1	DSPP_HUMAN
20	169	3.7	1389	1	DC13_DROME
21	168.5	3.7	841	1	IEG3_MCMVS
22	168	3.7	910	1	DNJM_MYCPN
23	168	3.7	527	1	ATRX_MOUSE
24	167.5	3.7	704	1	YH6_YEAST
25	167.5	3.7	1140	1	YH6_YEAST
26	166.5	3.7	1359	1	PCGV_HUMAN
27	166.5	3.7	3396	1	PCGV_HUMAN
28	166.5	3.7	1359	1	PCGV_HUMAN
29	165.5	3.7	1210	1	AF4_HUMAN
30	165.5	3.7	1435	1	EB41_PLAFC
31	165.5	3.7	3924	1	ANK2_MOUSE
32	164.5	3.6	1419	1	AL41_CANAL
33	163	3.6	1466	1	SPA2_YEAST

34	162.5	3.6	699	1	NP14_HUMAN
35	161.5	3.6	671	1	CHS5_YEAST
36	160.5	3.5	1356	1	HET1_PODAN
37	160.5	3.5	4377	1	ANK2_HUMAN
38	160	3.5	480	1	TGN2_HUMAN
39	159.5	3.5	1447	1	BUD4_YEAST
40	159	3.5	1275	1	YAU9_SCHPO
41	159	3.5	1310	1	YB35_SCHPO
42	158	3.5	2774	1	MAPA_RAT
43	157.5	3.5	680	1	NASP_RABIT
44	157.5	3.5	1912	1	VIT1_CHICK
45	157.5	3.5	2842	1	APC_RAT
46	156.5	3.5	713	1	TUP1_YEAST
47	156	3.4	1015	1	HIRA_MOUSE
48	156	3.4	1070	1	PVDG_PLAKN
49	155.5	3.4	1020	1	NPH_HUMAN
50	155.5	3.4	1217	1	AF4_MOUSE
51	155	3.4	687	1	DSPP_RAT
52	155	3.4	1062	1	SUM1_YEAST
53	154.5	3.4	589	1	HIBN_XENLA
54	154.5	3.4	1095	1	NEB1_RAT
55	154	3.4	1093	1	TMF1_HUMAN
56	154	3.4	2476	1	ATRX_MOUSE
57	153	3.4	725	1	AGAI_YEAST
58	153	3.4	1181	1	NKX1_RAT
59	152.5	3.4	1781	1	AKAC_HUMAN
60	152.5	3.4	1839	1	CYAA_SACKL
61	152	3.4	623	1	HPC2_YEAST
62	151.5	3.3	601	1	YFK5_SCHPO
63	151.5	3.3	1139	1	HMW1_MYCGE
64	151.5	3.3	1403	1	YDF3_SCHPO
65	151.5	3.3	2440	1	NCRI_HUMAN
66	151	3.3	572	1	CAF2_MOUSE
67	151	3.3	1075	1	MSB2_YEAST
68	151	3.3	1306	1	FL05_YEAST
69	150.5	3.3	943	1	CENC_HUMAN
70	150.5	3.3	1233	1	YF16_YEAST
71	150.5	3.3	1972	1	P531_HUMAN
72	150.5	3.3	2453	1	NCRI_MOUSE
73	150	3.3	340	1	GBB1_HOMAM
74	150	3.3	651	1	CORO_YEAST
75	150	3.3	679	1	YIS3_YEAST
76	150	3.3	742	1	CD44_HUMAN
77	149.5	3.3	1322	1	YAG3_YEAST
78	149.5	3.3	1585	1	P3K3_DICDI
79	149.5	3.3	1805	1	NEST_RAT
80	149	3.3	1155	1	PVDB_PLAKN
81	149	3.3	1162	1	TCNA_TRYCR
82	148.5	3.3	562	1	CHS5_CANAL
83	148.5	3.3	633	1	MLH_TETTH
84	148.5	3.3	915	1	NFM_HUMAN
85	148.5	3.3	1229	1	NP121_HUMAN
86	148.5	3.3	1637	1	MRSP_STAAU
87	148.5	3.3	1664	1	SLPI_CLOTM
88	148.5	3.3	2349	1	MSPI_PLAYO
89	148.5	3.3	2805	1	MAPA_HUMAN
90	148.5	3.3	2805	1	YF72_CAEEL
91	148	3.3	2132	1	PGCA_MOUSE
92	148	3.3	427	1	YKTS_YEAST
93	147.5	3.3	856	1	Y249_HUMAN
94	147.5	3.3	2090	1	N214_HUMAN
95	147.5	3.3	2843	1	APC_HUMAN
96	147.5	3.3	3866	1	HRX_MOUSE
97	147.5	3.3	3969	1	HRX_HUMAN
98	147.5	3.3	375	1	PEX7_YEAST
99	147	3.2	377	1	GBB2_TORAC
100	147	3.2	620	1	CHS5_SCHPO
101	147	3.2	1142	1	PAK1_YEAST
102	147	3.2	1453	1	NKCR_MOUSE
103	147	3.2	1453	1	NCAL_CHICK
104	146.5	3.2	1411	1	Y297_HUMAN
105	146.5	3.2	1411	1	Y297_HUMAN
106	146	3.2	377	1	GBB_ARATH

Q14978 homo sapien  
Q12114 saccharomyc  
Q00808 podospora a  
Q12955 homo sapien  
P47136 saccharomyc  
Q10164 schizosacch  
Q14340 schizosacch  
P34926 rattus norv  
P27123 oryctolagus  
P87498 gallus gall  
P70478 rattus norv  
P16646 mus musculu  
P50494 plasmodium  
P12036 homo sapien  
Q88573 mus musculu  
Q62598 rattus norv  
P46676 saccharomyc  
P66180 xenopus lae  
Q35867 rattus norv  
P82094 homo sapien  
Q61687 mus musculu  
P32323 saccharomyc  
Q92266 rattus norv  
Q02952 homo sapien  
P23466 saccharomyc  
Q01448 saccharomyc  
P87132 schizosacch  
Q49413 mycoplasma  
Q10475 schizosacch  
Q75376 homo sapien  
Q94077 mus musculu  
P38894 saccharomyc  
P23334 saccharomyc  
Q03188 homo sapien  
P43597 saccharomyc  
Q12888 homo sapien  
Q60974 mus musculu  
Q45040 homarus ame  
Q06440 saccharomyc  
P05653 saccharomyc  
P16070 h cd44 anti  
P39712 saccharomyc  
P54675 dictyosteli  
P21263 rattus norv  
P50493 plasmodium  
P23253 trypanosoma  
Q74161 candida alb  
Q40631 tetrahymena  
Q97293 homo sapien  
P80544 staphylococ  
Q06852 clostridium  
P13828 plasmodium  
P12270 homo sapien  
P78559 homo sapien  
Q92221 caenorhabdi  
Q61282 mus musculu  
P36046 saccharomyc  
Q92539 homo sapien  
P35658 homo sapien  
P25034 homo sapien  
P55200 mus musculu  
Q03164 homo sapien  
P39108 saccharomyc  
P93398 nicotiana t  
Q92357 schizosacch  
P38990 saccharomyc  
P30415 mus musculu  
P13590 gallus gall  
Q15040 homo sapien  
P49177 arabidopsis

Wed Mar 26 09:38:55 2003

340 SLQMSDMLSRFEAEVAQ---SNRGRSRPRGTSQSDISTLPTVPSSPDLEVSE 396  
145 SKDHSDEADR--PEAGDSTQSESEYRVGGSESSHGDCS-----EFDD 191  
397 TAMEVDTPABQFLQPTSTSTMSQAQHSSTSP-----ESPSTPLSSPDSEORQSVAS 451  
192 EGMQSDDP-----GSTRSDRGHTRMSAGIRSEESKGDHEPTSTQSDSDSDVEFS 242  
452 GHVTHOSDNNKLSKPGTGPVLSLHYSTGTSTIKLNFDEWSSIASRSGIGS 511  
243 SRKSFRRS-----RVSEDDRGELADNSRSTQSVSTDFRSK-----E 281  
512 HCKSGQSESVPOSSVQPPG-----DSETKAPEESEDVTKYQGYSAENPVENHINI 566  
282 ERSSTQEDTAETQSDQSDPEGQPSSESEAEAGEPSQESSESGVASESRGNDPNT 341  
567 TQSDKFTAKPLDSNGERDNLDRSCGVPPEASSEKAKPETSD-----QTSSTSATN 621  
342 SQ---TGQDRDSESEEDRLNTFSS-----SESQSTEEQSDSESNLSLSEESQESAQD 393  
622 ENNTNPEP-QFQTEATGPFSAHEETSTRDSALQDITDSD 658  
394 EDSSESQEGIQSQASRESRSQESQSDSRSENRDSD 431

RESULT 2

DMPL\_MOUSE  
ID DMPL\_MOUSE STANDARD; PRT; 503 AA.  
AC O55188;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Dentin matrix acidic phosphoprotein 1 precursor (Dentin matrix  
protein-1) (DMP-1) (AG1).  
DE protein-1 (DMP-1) (AG1).  
GN DMP1 OR DMP.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-Swiss Webster; TISSUE-Molar;  
RC MEDLINE=98184255; PubMed=9525343;  
RA McDougall M., Gu T.T., Luan X., Simmons D., Chen J.;  
RT "Identification of a novel isoform of mouse dentin matrix protein 1:  
spatial expression in mineralized tissues.";  
RL J. Bone Miner. Res. 13:422-431(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=129/SVJ; TISSUE=Spleen;  
RC Feng J.Q., Traianedes K., Luan X., McDougall M.;  
RT "Study of murine Dmp-1 gene function and regulation";  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: COULD BE INVOLVED IN THE MINERALIZATION OF EXTRACELLULAR  
MATRIX.  
CC -1- TISSUE SPECIFICITY: EXPRESSED IN TOOTH PARTICULARLY IN  
ODONTOBLAST, AMELOBLAST AND CEMENTOBLAST. ALSO EXPRESSED IN BONE  
PARTICULARLY IN OSTEOBLAST.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; U65020; AAB93764.1;  
CC EMBL; AJ242625; CAB59629.1;  
CC MG; MG1:94910; Dmp1.  
CC Extracellular matrix; Signal.  
CC POTENTIAL.  
CC SIGNAL 1 503  
CC CHAIN 17 503 DENTIN MATRIX ACIDIC PHOSPHOPROTEIN 1.

P19246 mus musculus  
P32768 saccharomyc  
Q93318 drosophila  
P82179 canis famil  
P38904 saccharomyc  
Q9ulh0 drosophila  
P93563 solanum tub  
O09715 schizosacch  
P16170 xenopus lae  
O61315 mus musculus  
Q09950 caenorhabdi  
Q23256 caenorhabdi  
P14737 saccharomyc  
O60309 homo sapien

ALIGNMENTS

RESULT 1  
DMPL\_MOUSE STANDARD; PRT; 489 AA.  
AC P98193;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Dentin matrix acidic phosphoprotein 1 precursor (Dentin matrix  
protein-1) (DMP-1) (AG1).  
GN DMPL.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-Sprague-Dawley; TISSUE-Tooth;  
RC MEDLINE=93286101; PubMed=8509401;  
RA George A., Sabsay B., Simonian P.A., Veis A.;  
RT "Characterization of a novel dentin matrix acidic phosphoprotein.  
Implications for induction of biomineralization.";  
RL J. Biol. Chem. 268:12624-12630(1993).  
CC -1- FUNCTION: COULD BE INVOLVED IN THE INDUCTION OF MINERALIZATION OF  
EXTRACELLULAR MATRIX.  
CC -1- TISSUE SPECIFICITY: EXPRESSED IN TOOTH PARTICULARLY IN OODONTOBLAST  
AND AMELOBLAST.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; L11354; ; NOT\_ANNOTATED\_CDS.  
CC Extracellular matrix; Signal.  
CC SIGNAL 1 16 POTENTIAL.  
CC CHAIN 17 489 DENTIN MATRIX ACIDIC PHOSPHOPROTEIN 1.  
FT DOMAIN 41 44 POLY-PRO.  
FT CARBOHYD 340 340 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 378 378 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 443 443 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT SITE 334 336 CELL ATTACHMENT SITE (POTENTIAL).  
CC SEQUENCE 489 AA; 53058 MW; 59F8381479DDA085 CRC64;  
Query Match 4.9%; Score 221.5; DB 1; Length 489;  
Best Local Similarity 25.9%; Pred. No. 0.00018;  
Matches 103; Conservative 56; Mismatches 164; Indels 75; Gaps 14;

QY 281 DPKDDT-AREUKTPAERRELQPPVKRLRLRGDSMTGTPRPRPSERERDGEQSPNV 339  
DB 89 DSGDDTDFGDDNGPGPEER---QWGGPSRLDSDSDSDTQSSDSTSQNSAQDTPSD 144



Query Match

Matches 105; Conservative 58; Mismatches 153; Indels 160; Gaps

291 KTPSAEERREELR-QPPVKRLRLRGDWSDTGPRARPESEERDGEQSPNVSLMQRMSDML 349

350    S A W F E A S V A Q S N R G R S R P R G T S Q ----- 377  
               ||: || | | | :  
78 -----

SDISTLPTVPSSPDLEVTAMETPAEQFLQPS----- 412

123 OWGGSPKIKNSDENSADMMGGSTG----- 412

```

182 SESEORVGGSGEGSSGCGCSCSNNNGCCGCGC-----TSSSTMSAQAHSTSSPT-----ESPHST 434
      ||| | : |

```

242 EPTSTQSDSDSQSVFSSRSKSFRRS-----PVCTGCTGGTCT  
| : ||: |||| | : | : : | ||  
| : ||: |||| | : | : : | ||

[illegible]

```

|||||::: | | : |
341 SQEGVTSERGDNPTISQ---TGQDESESESSEEDSINTFSS-----SPSGCSMBBPSAAS-

```

[illegible]

STANDARD.	SR40 YEAST	_YEAST
100	100	100
200	200	200
300	300	300
400	400	400
500	500	500
600	600	600
700	700	700
800	800	800
900	900	900
1000	1000	1000

01-JUN-1994	(Rel. 29, Last sequence update)
15-JUL-1998	(Rel. 36, Last annotated sequence)

Saccharomycetes cerevisiae (Baker's yeast).  
Eukaryota; Fungi; Ascomycota: Saccharomycetia.

[1]  
SEQUENCE FROM N.A.

[2] submitted (MAY-1993) to the EMBL/GenBank/DBJ databases.

Sou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G., Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta I.

\_\_\_\_\_

Wed Mar 26 09:38:55 2003

RA Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D., Churche C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A., Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M., Louis E., Iye G., Moule S., Mouton T., Odell C., Pearson D., Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V., Walsh S.V., Whitehead S., J. Bacteriol. 169:2142-2149 (1987).

RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE OF 1-242 AND 762-1331 FROM N.A.

RX MEDLINE-87194600; PubMed-3106330;

RA Yamashita I., Nakamura M., Fukui S.;

RT "Gene fusion is a possible mechanism underlying the evolution of

RT STAL.";

RL J. Bacteriol. 169:2142-2149 (1987).

RN [3]

RP SEQUENCE OF 1-31 FROM N.A.

RX STRAIN-SPX101-1C; PubMed-3141213;

RA MEDLINE-89031230; PubMed-3141213;

RT Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;

RT "Similar short elements in the 5' regions of the STA2 and SGA genes

RT from Saccharomyces cerevisiae.";

RL FEBS Lett. 239:179-184 (1988).

CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-

CC glucose residues successively from non-reducing ends of the chains

CC with release of beta-D-glucose.

CC -1- SIMILARITY: TO S.POMBE SPBC215.13.

CC -1- SIMILARITY: SOME. TO S.POMBE SPCC285.13C.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

DR EMBL; Z38061; CAA86176.1; -

DR EMBL; M16164; AAA35014.1; -

DR EMBL; M16165; AAA35015.1; -

DR EMBL; X13857; CAA32069.1; -

DR PIR; B26877; B26877.

DR PIR; A26877; A26877.

DR PIR; S48478; S48478.

DR SGD; S0001458; MUC1.

DR Hydrolyase; Glycosidase; Polysaccharide degradation; Glycoprotein;

KW Signal; Multigene family.

FT SIGNAL 1 21 POTENTIAL.

FT CHAIN 22 1367 GLUCOAMYLASE S1/S2.

FT DOMAIN 210 1367 SER/THR-RICH.

FT CARBOHYD 817 817 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 874 874 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 1367 AA; 136110 MW; 91C00E2BD61AA9D CRC64;

Query Match 4.2%; Score 190.5; DB 1; Length 1367;

Best Local Similarity 22.3%; Pred. No. 0.021;

Matches 124; Conservative 64; Mismatches 270; Indels 99; Gaps 18;

QY 151 NDRPVTFLSCGSDGVNRFDFTRKTSCTKEDKDDILNCRAATSAVCAIPPIYLAAG 210

DB 272 PPTPTTCTKTKPPPHDIT---TPCFK--KTTSTKCTKTKTT-----PVPTPSSST 321

QY 211 CSDSSVRIYDRMLTRATGAVAGTGTGMVARETP-----SHLNKSKCRVTSACYSEDGQ 266

DB 322 TESSAPVPTPSSSTTESSAPVPTSSSTTESSAPVPTSSSTTESSAPVPTSSSTTESSA 381

QY 267 EILVSVSSDIYLEDKDDTARELTPSAERREELRPPVKRLRLGWDGTPRARP 326

DB 382 PVTSST-----ESSAPVPTPS-----STESSAPV 410

QY 327 SERERDGEOSPNNVLMQRMDSMLSRWFEEASVAQSNCRGSRP--RGTSQSDISTLPT 385

DB 411 TSSTTESSAP-----VTSSTTESSAPVPTSSSTTESSAPVPTSSSTTESSAPVPT 460

QY 386 VPSPDLEWSE---TAMEVDTPAEQFLQPTSTSMQAQAHSTSPTES---PHSTPLLS 439

DB 461 -PSSSTTESSAPVPTSSSTTESSAPVPTSSSTTESSAPVPTSSSTTESSAPVPTSSST 519

QY 440 PDSQROVSASGHHTHQS-----DNNNEKLS---PKPTGEPVLSLHYSTEGTSTIK 492

DB 520 TESSAPAPVPTSSSTTESSAPVPTSSSTTESSAPVPTSSSTTESSAPVPTSSSTTESSA 579

QY 493 LNFTDEWSSIASSRGIGSCHCKSGQESF---VPQSSVQPPGDSGSETKAPESSESDVK 549

DB 580 PVPTPSSSTTESSAPVPTSSSTTESSAPVPTSSSTTESSAPVPTSSSTTESSAPVPT 639

QY 550 YQEGVSAEN---PVENHINITQSKFTAKPLDSNGERNDLNRSCGVPEASASSEKAK 606

DB 640 TPSSSTTESSAPVPTSSSTTESSAPVPTSSS-----TTSSAPVPTSSSTSS 692

QY 607 EPETSDQSTESA-----TNENTNPEQFOFOTEA-----GPSAHEETSTRDSA 650

DB 693 APVTSSTTESSAPVPTSSSTTESSAPVPTSSSTTESSAPVPTSSSTTESSAPVPT 752

QY 651 LQDTDDSDDDPVLIPGA 667

DB 753 TSSTTESSAPVPTPSS 769

RESULT 5

MGC1\_HUMAN STANDARD; PRT; 1142 AA.

ID MGC1\_HUMAN

AC 060732; O75451;

DT 15-OCT-2001 (Rel. 40, Created)

DT 15-OCT-2001 (Rel. 40, Last sequence update)

DT 13-JUN-2002 (Rel. 41, Last annotation update)

DE Melanoma-associated antigen C1 (MAGE-C1 antigen) (Cancer-testis

DE antigen CT7).

GN MAGEC1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.

OX NCBL\_TaxID=9606;

RP [1]

RN SEQUENCE FROM N.A.

RC TISSUE-Melanoma;

RX MEDLINE-98143561; PubMed-9485030;

RA Lucas S., De Smet C., Arden K.C., Viars C.S., Lethe B., Lurquin C.,

RA Boon T.;

RA "Identification of a new MAGE gene with tumor-specific expression by

RA representational difference analysis.";

RT Cancer Res. 58:743-752(1998).

RL [2]

RN SEQUENCE FROM N.A.

RP TISSUE-Melanoma;

RX MEDLINE-98284033; PubMed-9618514;

RA Chen Y.-T., Gure A.O., Tsang S., Stockert E., Jager E., Knuth A.,

RA Old L.J.;

RA "Identification of multiple cancer/testis antigens by allogeneic

RA antibody screening of a melanoma cell line library.";

RT Proc. Natl. Acad. Sci. U.S.A. 95:6919-6923(1998).

RL [3]

RN SEQUENCE FROM N.A.

RP Bird C., Heath P.;

RA Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.

CC -1- TISSUE SPECIFICITY: Expressed in testis and in tumors of a wide

CC variety of histologic types.

CC -1- SIMILARITY: CONTAINS 1 MAGE DOMAIN.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

P45384;  
01-NOV-1995 (Rel. 32, Created)  
01-NOV-1995 (Rel. 32, Last sequence update)  
15-JUN-2002 (Rel. 41, Last annotation update)  
Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGAL protease).  
IGA.  
Haemophilus influenzae.  
Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
Haemophilus.  
NCBI\_TaxID=727;  
[1]  
SEQUENCE FROM N.A.  
RP STRAIN=HK715 / Serotype B;  
RX MEDLINE=92234949; PubMed=1373717;  
RA Poulsen K., Reinholdt J., Killian M.;  
RT "A comparative genetic study of serologically distinct Haemophilus  
influenzae type 1 immunoglobulin A1 proteases.";  
RL J. Bacteriol. 174:2913-2921(1992).  
CC -!- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A  
CC PRODUCING INTACT FC AND FAB FRAGMENTS.  
CC -!- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at  
CC certain Pro-I-Xaa bonds in the hinge region. No small molecule  
CC substrates are known.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC  
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE  
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE  
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY  
CC SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.  
-----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; M87489; AAA34966.1; -  
DR MEROPS; S06.001; -  
DR InterPro; IPR000710; IGA\_S6.  
DR InterPro; IPR004899; Pertact\_sup.  
DR Pfam; PF02395; IGA1; 1.  
DR Pfam; PF03212; Pertactin; 2.  
DR PRINTS; PR00921; IGASERPTASE.  
DR HydroLase; Serine protease; Transmembrane; Zymogen; Repeat; Signal.  
KW SIGNAL 1 25  
FT CHAIN 26 1014  
FT PROPEP 1015 1702  
FT ACT\_SITE 288 288  
FT DOMAIN 1109 1124  
FT REPEAT 1109 1116  
FT REPEAT 1117 1124  
FT SEQUENCE 1702 AA; 186539 MW; 860F70D2667807A6 CRC64;  
-----  
Query Match 4.28; Score 188; DB 1; Length 1702;  
Best Local Similarity 18.98; Pred. No. 0.036;  
Matches 150; Conservative 120; Mismatches 311; Indels 212; Gaps 33;  
-----  
67 LSGDDTKVLISNPYSRKVLTTIRSGHRANIFSAKFLPCTNDKQIVYSCSDGVIFYTNVE 126  
773 ITASDIAKAVHG--YKAGDTVCVRSDDTG-----YTCITDK--LSDKALNSFNATNVS 822  
127 QDAETNRQCQFTCH-----YCT-----TYEIMTVPNPDYFLSCGEDGTGRWRFTRIKTS 177  
823 GNVNLSGANFVLGKANLFTISGTNSQVRLTENSHWHLTG---DSNVQLNLD----- 874  
178 TKEDCKDDILINCRAATSVIAICPPIP-----YYLAVGCSDDSSVRIYDRMLGTR 227  
875 -----KGHHLNQAQNDANKVTYNTLVNLSNGSFFYL-----TDLNKGDKGVVVTKS 925

Wed Mar 26 09:38:55 2003

```

QY 228 ATGNY-----AGRTTGMVARFIPSHLNKSKRVTSCLYSED---GQELVSYSDYI 277
Db 926 ATGNTLQVADTKGTPKNELTLDASNATRNNLNSLVGNTVDLGAWKYKLRNVNGRY- 984
QY 278 YLFDPK-----DTPAR-----ELKTPSAERRELRQ-----PPVKRLRLRGDWS 318
Db 985 DLYNPEVEKRNQTDVDTNITPNNOADVPSPNSNEELARVETVPVPPAPATPSETTET 1044
QY 319 TGPRAPE-----ERERDQ-----EQSPNVSLMQRMSDMLSRWFE-----354
Db 1045 VAENQKSEKTEKNEKQDQATEITTAQNGEVAEAKPSKANTQNTNEAQSGETEETOTTE 1104
QY 355 --BASEVAQNRGRGRSRPGTQSQSDISLTPVPS-----SPDLEVSETA 398
Db 1105 IKETAKVEEKEKARKEKAKVEKEIQAPQAMASETPKQAKPAPKEVSTDTKVEEQ 1164
QY 399 MEVDTPAEQFLOPSTSTMSAQAHSTSPSTESPHSTPLSSPDSEORQSEAGSHHTHQ 458
Db 1165 VQA-----QPQTSTVAAAEAT-----SPNSPAEETQPSKTAEPVTPVVKNQ 1211
QY 459 SDNNEKLSKPGTGPVLSHYSTEGTITSTIKLNFDEWSSSIASSRGIGSHCKSEGO 518
Db 1212 TENT-----TDQP-----TEREKAKVETKTEKTPQPVASQA-----SPKQEQ 1249
QY 519 BESFVQSSVOPPEGSDSEKAPESSESDVTKYQGVSAENPV-ENHIN-----ITQSDK 571
Db 1250 SETVQPAVLSENVPTVYNAEEVQAOLQOTTSATVSTKQAPAPENSINTGSATAITAE 1309
QY 572 FTAKPLDSNGERDNLNDRSCGVPEESASSEKAKEPETSQDSTESATENNTNPEPF 631
Db 1310 KSDKP-----QTETAASTEDASQHKANTVADNSVANNSESSEPKSRR 1351
QY 632 QTEATGSAHEETSTDSALQDSDSDVLPICGARYAGGDRRSARVARIQEFFRRK 691
Db 1352 RRSISQ---QETSARETTAASTDEI-----TIADNSKRSP-NRRS-----RRS 1392
QY 692 ERKEMBELDTLNRRLPLVMKYKGRHSRTMIKEANFWGNCVMSGDCGHIFI---WDR 748
Db 1393 VRSE-----PTVTNGSRSTVALRDLTSTNTNAVISDAMAKAQFVALNVGK 1438
QY 749 HTAEHMLLEADN 761
Db 1439 AVSQHISQLEMN 1451

RESULT 7
ID IGAO_HAEIN STANDARD; PRT; 1694 AA.
AC P44969;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGAL protease).
GN IGA OR IGAI OR HI0990.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Serotype D;
RA Wright A., Fishman Y., Tai F., Plaut A.G.;
RL Submitted (MAY-1991) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA Mckenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,

```

```

RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RD Rd.";
RL Science 269:496-512(1995).
CC -1- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC PRODUCING INTACT FC AND FAB FRAGMENTS.
CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC certain Pro|-xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; X59800; ; NOT_ANNOTATED_CDS.
CC EMBL; U32779; AAC22651.1; -.
CC MEROPS: S06.001; -.
CC TIGR: HI0990; -.
CC InterPro: IPR000710; IGA_S6.
CC Pfam: PF02395; IGAI; 1.
CC Pfam: PF03212; Pertactin; 2.
CC PRINTS: PR00921; IGASERPTASE.
CC Hydrolase; Serine protease; Transmembrane; Zymogen; Signal;
CC Complete proteome.
CC SIGNAL 1 25
CC CHAIN 26 1014
CC PROPEP 1015 1694
CC FT ACT_SITE 288
CC FT CONFLICT 253
CC FT CONFLICT 272
CC FT CONFLICT 464
CC FT CONFLICT 866
CC FT CONFLICT 1036
CC FT CONFLICT 1074
CC FT CONFLICT 1421
CC FT CONFLICT 1545
CC FT CONFLICT 1694 AA; 185539 MW; C52427013F93178C CRC64;
SQ SEQUENCE
Query Match 4.1%; Score 186; DB 1; Length 1694;
Best Local Similarity 19.4%; Pred. No. 0.045;
Matches 153; Conservative 113; Mismatches 310; Indels 214; Gaps 33;
QY 67 LSGSDDTKLVISNPYSRKVLTITIRSHRANIFSAKLPCTNDKQIVSCSDGVVFTNVE 126
Db 773 ITASDNKAVHIG--YKAGDTVCVRSDYTG-----YVCTTDR--LSDKALNSFNATNVS 822
QY 127 QDAETNRQCOFTCH-----YGT-----TYEIMTVPNPDYFLSCGEDGTVRWFDTIKTSC 177
Db 823 GNVNLGNANFVGLKANLFGTISGTGNSQVRLTENSHTLG---DSNVNQLNLD-----874
QY 178 TKEDCKDDILNCRRAATVAICPPIP-----YLLAVGSDSSSVRIYDRMLGTR 227
Db 875 -----KGHIHLNAQNDANKVTTYNTLVNSLSGNGSFYLL-----TDLNKGQGVVTKS 925
QY 228 ATGNY-----AGRTTGMVARFIPSHLNKSKRVTSCLYSED---GQELVSYSDYI 277
Db 926 ATGNTLQVADTKGTPKNELTLDASNATRNNLNSLVGNTVDLGAWKYKLRNVNGRY- 984
QY 278 YLFDPK-----DTPAR-----ELKTPSAERRELRQ-----PPVKRLRLRGDWS 318

```

```

Db 985 DYNPEVKRKNQVDTNITTTNNIQADVPSVPSNNEETARVETPPPPAPAPPSETTET 1044
QY 319 TGPARPSEPERERDGEQSPNVSLMORMSDMLSRWFEEA-----SEVAQS----- 362
Db 1045 VAENSKOESKTVKKNQDQATTAQN-----GEVAEEAKPSVKANTQTNEVAQSGSETEE 1099
QY 363 -----NRGRSRPRGGTSGDSTLTPTVPS-----SPDLEVSAMEV 401
Db 1100 TOTTEIKETAKVEKAKVEKDEIQEAPOMASETSPKQAKPAKEVSTDRKVEETQVOA 1159
QY 402 DTPAEQFLQPTSTSTMSAQAHSTSPSPHSPTLLSSPDSEQSQSVBASGHHHTHQSDN 461
Db 1160 -----QPQTQSTTVAAAEAT-----SPNSKPAEETQPSSEKNAEPVTPVSKNQEN 1206
QY 462 NNEKLSRPGGEPVLSHYSTEGTITIKLNFDTDEWSSIASSSRGIGSHCKSEGQES 521
Db 1207 T-----TDQP-----TEREKTAKVETKTEQPPQVASOA-----SPKOEQSET 1244
QY 522 FVPOSSVQPPGEGSETKAPESSEDVTKYQGVSAENPV-ENHIN-----ITQSDKFTA 574
Db 1245 VQPAVLESENVPVNNAEVQALQQTTSATVSTKQAPENSINTGSATAITETAESD 1304
QY 575 KPLDSNGERDNLDRSCGVPEESASSEKAKPETSDQTSATNTENNTPQOTE 634
Db 1305 KP-----QTETAATEDASQHKANTVADNSVANNSESDPKRRRRS 1346
QY 635 ATGPSAHEETSTRDSALQDQDSDVPLIPGARYAGDRSARVARIQEFFRRK 694
Db 1347 ISQP-----QETSARETTAASTDET-----TIADNSKRSKP-NRRS-----RSVRS 1387
QY 695 EMEELTLNIRPLVKWYKGRHNSRTMIKEANFWANFVMSGDCGHFI---WDRHTA 751
Db 1388 E-----PTVNGSDRSTVALRDLTSTNTNVAISDAMAKAQFVALNWKAYS 1433
QY 752 EHLMLLEADN 761
Db 1434 QHISOLEMNN 1443

```

## RESULT 8

```

HIRA_FUGRU
ID HIRA_FUGRU STANDARD; PRO: 1025 AA.
AC 042611;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE HIRA protein (Tup1 like enhancer of split protein 1).
GN HIRA OR TUPL1.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
OX NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98201624; PubMed=9524281;
RA Llevadot R., Estivill X., Scambler P., Pritchard M.;
RT "Isolation and genomic characterization of the TUPL1/HIRA gene of
RL the pufferfish Fugu rubripes."
RL Gene 208:279-283(1998).
CC -!- FUNCTION: Could have a part in mechanisms of transcriptional
CC regulation similar to that played by yeast HIR1 and HIR2 together.
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- SIMILARITY: CONTAINS 4 WD REPEATS (TRP-ASP DOMAINS).
CC -!- SIMILARITY: BELONGS TO THE HIR1 FAMILY OF WD-REPEAT PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U94325; AAC60370.1; -
DR EMBL; U94324; AAC60369.1; -
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40. 8
DR PRINTS; PR00320; GPROTEINBRPT.
DR ProDom; PD000018; WD40. 3.
DR SMART; SM00320; WD40. 5.
DR PROSITE; PS00678; WD_REPEATS_1; 1.
DR PROSITE; PS00082; WD_REPEATS_2; 3.
DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
KW Transcription regulation; Repeat; WD repeat; Nuclear protein.
FT REPEAT 68 98
FT REPEAT 129 159
FT REPEAT 172 202
FT REPEAT 266 313
FT DOMAIN 673 682
FT DOMAIN 685 688
FT POLY-ALA.
SQ SEQUENCE 1025 AA; 111856 MW; A4212152D75B6A37 CRC64;

Query Match
Best Local Similarity 4.13; Score 184; DB 1; Length 1025;
Matches 175; Conservative 116; Mismatches 312; Indels 272; Gaps 39;

QY 42 KLEATLVNHDGCVNTICWNTDTEVILSGSDDTKLV-----ISNPY 81
Db 61 KMLCOMDNHLACVNCVNSNGLYASGGD-KLVVWVKKRAALIGPSTVFSSKNLANVE 119
QY 82 SRKVLVTIRSHRANTFSAKFLPCTNDKQIVSCGDCGVIFVTNVQDAETNRQCFCHY 141
Db 120 QWRCVTLIRN-HTGDVMDYSWSP--HDVWLASCSVDNTIIVNARKEPE-----MTCLR 171
QY 142 GTTYEIMTVNDPY--TFLSCGEDGTVR-W-FDTRIKTSCTKEDCKDDILINCRRAATS 196
Db 172 GHTGLVKGLTWDVPVKYIASQADDSLRVWRTVDMWMEANITKPFSECGGTTHVLRLSWS 231
QY 197 VAICPPPIPYL-----AVGSDSSVRIYDRMLGTRATGNAGRTTGMVARETPSHLNK 252
Db 232 -----PDGQYLVSAHAMNNSGPTAQIVERD--GWRTNMDVFGHRKAVTVVKNPKFKKK 284
QY 253 -----SC-----RVTSLSYSEDEQEL 269
Db 285 QKNGGSPKPCYCCCAVCGSKDRSLVMTSLKRLPLVWIHDLFKSIMDISWTLTGLGML 344
QY 270 V-SYSSDYIYLEDPKDDTARELKTPSAEERRELRQ-----P 305
Db 345 VCSMDGTVAYL-----DFSLDELGDPLSEEEKNSIHQNIYKSLAITNTEPQLSTIENP 400
QY 306 PV---KRLRLRGDWSDTGP-----RAPERERDRDQSPNVSLMORMSDMLSRWFEEA 356
Db 401 EMLKYQOERRNSTQANSQPGATGESATPKLNSVMNGE-----SLEDIRKNLLKQVE-- 453
QY 357 SEVAQSNRGRSRP-----RGTSQSDISTLTPTVPSPPDLEVSETAMEVDTPAEQFL 409
Db 454 ---TRTPDGRRIPTPLCIAQLDGTDFSPALFNSAPILPSG----- 490
QY 410 OPSTSTMSAQAHSTSPSTES-----PHSTPLSSPDSEQSQSVBASGHHHTHQSDNNN 463
Db 491 -SSMSNLTQLSSDSSPGQAPPLGLRPSQDPLMISPPSSAAKV-----LEDNKD 540
QY 464 -----EKLSPKPGTGPVLSHYSTEGTITIKLNFDTDEWSSIASSSRGIGSH 512
Db 541 GVKSCILLTSASKIEPMKALDSRFTERSKATPGAT-----AATASST-GL-- 584
QY 513 CKSEGESEGFVPOSSVQPPPE---GDSETK-----APE---ESSEVDVTKYOE 553
Db 585 TSPERKESTPKQDKVEDTSSDESKMATINKNLFNKRKPELLMDGAEEVKKKG 644
QY 554 VSAENPVENHINTOSDKFTAKPL--DSNSGENDLNLDRSCGVPEESASSEKAKPET 611
Db 645 RPRKDMAASI-----AQPLTQTTSPAREP---SRAAAGAGAAAPTAAAKLP 692

```

QY 486 T-TTSTIKLNFDE-----WSSIASRRGIGSHCKSEGEFVPPQSSVQPPGDSSET 537  
Db 355 AFDFGNSVNSVDSEDEESQAESYSAENAVYHNEHEDLDKELIEDIE-----SSDSES 410  
QY 538 KAPEES---SEDVTKVQGVSAENPVENHINITQSKFTAKPLDSSNGERNOLNDRSCG 594  
Db 411 QSAQESQSGSDDEPEYK-----MKNEKSTSEETE-NISESRDOGFADKAY 454  
QY 595 VPESASSEKAKPETSDQSTESATNENNTNPEQFQTEA-----TGPSAHEETSTRDSA 650  
Db 455 TKNVQEQEENDEPEKDDIIRSLDKNFHGNKSEYSENVLENETDPAIVE-----RENQ 510  
QY 651 LQDTDDSDDDPVLIPGARYAGPDRRSARVARIQEFFRRKE-----KEMEELDTLNR 706  
Db 511 INDVEGYDVTGKSVESDLHEHSPDNLXDLAARMLQFQSSNSCPOKEQ-----561  
QY 707 PLYKMYKGRN-----SRTMIKEANFWANFVMSGDCGHIFWDRITAHLM 755  
Db 562 --VSESYLGHNSGNSLSRSLDESEEQIPLKDFTCENNNLTKDRGDLSSSVEIEVEKVS 619  
QY 756 LLEADNHVYVNCLOPHFPDPIILASSGIDYDIKIWSPLESRIFENRKLAD-----EVT--RN 809  
Db 620 EKKLDGSTKELVPLSTDTTINNSSLGNEISYISLDDADAISEMLTDPVPLMEIKTFKY 679  
QY 810 ELMLEET-----RNTITVP-----ASFMLRLASLN 835  
Db 680 EYVISESVSYSTSYEDNTVAMPPOVEYTSFPFNDPFNSLN 719

RESULT 10

IGM4\_HAEIN STANDARD; PRT; 1849 AA.  
ID IG44\_HAEIN  
AC P45386; (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGAL protease).  
GN IGA.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus  
OC Haemophilus  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-NHTI HK61;  
RX MEDLINE=92234949; PubMed=1373717;  
RA Poulson K., Reinholdt J., Killian M.;  
RT "A comparative genetic study of serologically distinct Haemophilus  
influenzae type 1 immunoglobulin A1 proteases."  
RL J. Bacteriol. 174:2913-2921(1992).  
CC -1- FUNCTION: VIRULENCE FACTOR: CLEAVES HOST IMMUNOGLOBULIN A  
CC PRODUCING INTACT FC AND FAB FRAGMENTS  
CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at  
CC certain Pro-I-Xaa bonds in the hinge region. No small molecule  
CC substrates are known.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC  
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE  
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE  
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY  
CC SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; M87491; AAA24968.1;  
DR MEROPS; S06.001;

QY 612 DOTSTESATNENNTNPEQFQTEATGPSAHEETS-----TRDSALQDTDDSDDDPVLIPGA 667  
Db 693 TFSIKKAFTLQVSMDFSVVLEVE-----NEVSVAGSRLSQLRCSRDGRDWTLLPSS 745  
QY 668 RYRAGPDRRSAVA---RIQEFFRRKKEEMEELDTLNRPLVKKMYKGRHNSRTMIK 724  
Db 746 VITAAGSDVAVASQDRMLSVFSCGRL-----LPAIQLATP-----A 785  
QY 725 EANFWANFVMSGDCGHIFWDRHTAEHMLLEA 759  
Db 786 SALHCSAHEFVWLTSGATLSVMDVHKQKALVKNES 820

RESULT 9

YME7\_YEAST STANDARD; PRT; 1658 AA.  
AC Q03661; Q04988;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE Hypothetical 187.1 kDa protein in GUA1-ERG8 intergenic region.  
GN YMR219W OR YMR261.13 OR YW959.01.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE OF 1-711 FROM N.A.  
RC STRAIN=S288c / AB972;  
RA Dedman K., Brown D., Bowman S., Barrell B.G., Rajandream M.A.,  
RA Walsh S.V.;  
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 608-1648 FROM N.A.  
RC STRAIN=S288c / AB972;  
RA Skelton J., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;  
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; Z49809; CAA89934.1;  
DR EMBL; Z49939; CAA90190.1;  
DR SGD; S0004832; YMR219W.  
KW Hypothetical protein.  
SQ SEQUENCE 1658 AA; 187137 MW; 3893F968305A757D CRC64;  
Query Match 3.9%; Score 177.5; DB 1; Length 1658;  
Best Local Similarity 18.1%; Pred. No. 0.12;  
Matches 116; Conservative 110; Mismatches 243; Indels 171; Gaps 25;  
QY 278 YLEPDKDRTARELTPSAEERREL--RQPPVKRLRLRGDWSGTGPRPARESERDEGEQ 335  
Db 169 FISSPLPADESSNIDSKDEDELEGKQSLIKDFDLENDEYELS-----EEKNSDQGS 223  
QY 336 SPNVSL-----NORMSDMLSRWFEEASEVAOSNGRGRSRPRGDT--SQSDIST 382  
Db 224 SPSTIMLSDEYEAAGALQVSN--DEYAEEGQVGRKNIGQEQANVENATQISSDSE 281  
QY 383 LPTVPSPDLEVSAMEVDTPAE-----QFLQPTSTSTMSQAHAHST 425  
Db 282 GQNYSEGEMEL-EDDIDVSEDAEKDSQAGTEHSDVFSKYMPQRPDNT-----331  
QY 426 SPTEPHSTPLSSPDSEQRQSVASGHHTTHDSNNNEKLSPKPGTGPVLSLHYSTEG 485  
Db 332 -----KIPVIEKVESD-----HKVHQ-----RYSBDG 354

"Dentin phosphoprotein and dentin sialoprotein are cleavage products expressed from a single transcript coded by a gene on human chromosome 4. Dentin phosphoprotein DNA sequence determination.";

J. Biol. Chem. 272:835-842(1997).

[2]

SEQUENCE FROM N.A., AND REVISIONS TO C-TERMINUS.

RC STRAIN=129/SVJ; TISSUE=Liver;

RX MEDLINE=98211969; PubMed=9545272;

Feng J.Q., Luan X., Wallace J., Jing D., Ohshima T., Kulkarni A.B.,

RA DeSouza R.N., Kozak C.A., MacDougall M.;

\*Genomic organization, chromosomal mapping, and promoter analysis of

the mouse dentin sialophosphoprotein (Dspp) gene, which codes for both

dentin sialoprotein and dentin phosphoprotein.";

J. Biol. Chem. 273:9457-9464(1998).

[3]

SEQUENCE FROM N.A.

RC STRAIN=129/SVJ; TISSUE=Liver;

RFair C., Butler S., Lin E., George A., Veis A.;

\*From mouse to zebrafish-dentin matrix proteins genomic

characterization.";

Submitted (MAR-1999) to the EMBL/GenBank/DDJB databases.

[4]

TISSUE SPECIFICITY.

MEDLINE=98055479; PubMed=9395101;

Hatch H.H., Berry J.E., Somerman M.J., Bronckers A.L.,

Hotton D., Papageorgakis P., Berdal A., Butler W.T.;

"Dentin sialoprotein (DSP) transcripts: developmentally-sustained

expression in odontoblasts and transient expression in

pre-ameloblasts";

Eur. J. Oral Sci. 105:405-413(1997).

[5]

TISSUE SPECIFICITY.

MEDLINE=21096982; PubMed=11175790;

Xiao S., Yu C., Chou X., Yuan W., Wang Y., Bu L., Fu G., Qian M.,

Zhao J., Shi Y., Hu L., Han B., Wang Z., Huang W., Liu J., Chen Z.,

Zhang G., Kong X.;

"Dentogenesis imperfecta 1 with or without progressive hearing loss

is associated with distinct mutations in DSPP.";

Nat. Genet. 27:201-204(2001).

-|- FUNCTION: DSP may be an important factor in dentinogenesis. DPP

may bind high amount of calcium and facilitate initial

mineralization of dentin matrix collagen as well as regulate the

size and shape of the crystals.

-|- SUBCELLULAR LOCATION: Secreted (By similarity).

-|- TISSUE SPECIFICITY: Expressed in teeth, mainly in odontoblasts and

transiently in pre-ameloblasts. Found in the inner ear.

-|- PMM: DSP is glycosylated.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL Outstation-

The European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See http://www.isb-sib.ch/announce/

or send an email to license@isb-sib.ch).

-----

EMBL: U67916; AAC12787.1 ;

EMBL: AJ002141; CAU05208.1; ALT\_SEQ.

EMBL: AF135799; AAD42781.1; ALT\_SEQ.

GMD; MG1:109172; Dssp.

Signal; Extracellular matrix; Glycoprotein; Sialic acid;

Phosphorylation.

SIGNAL 1 17

FT CHAIN 18 934 POTENTIAL.

FT CHAIN 18 451 DENTIN SIALOPHOSPHOPROTEIN.

FT CHAIN 452 934 DENTIN SIALOPHOSPHOPROTEIN.

FT DOMAIN 419 934 SER/ASP-RICH.

FT SITE 479 481 CELL ATTACHMENT SITE (POTENTIAL).

FT MOD\_RES 227 227 PHOSPHORYLATION (BY CK2) (POTENTIAL).

FT MOD\_RES 254 254 PHOSPHORYLATION (BY CK1) (POTENTIAL).

FT MOD\_RES 279 279 PHOSPHORYLATION (BY CK1) (POTENTIAL).

FT MOD\_RES 293 293 PHOSPHORYLATION (BY CK2) (BY SIMILARITY).

FT MOD\_RES 299 299 PHOSPHORYLATION (BY CK1) (BY SIMILARITY).

-----

DSP\_MOUSE STANDARD; PRT: 934 AA.

15-JUN-2002 (Rel. 41, Created)

15-JUN-2002 (Rel. 41, Last sequence update)

Dentin sialophosphoprotein precursor (Dentin matrix protein-3) [DMP-

3] contains Dentin phosphoprotein (Dentin phosphoporphyn) (DPP);

Dentin sialoprotein (DSP)].

OC Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OX Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

RN NCBI\_Taxid=10090;

[1]

SEQUENCE FROM N.A.

RC STRAIN=Swiss Webster; TISSUE=Molar;

RX MEDLINE=97150835; PubMed=8995371;

MacDougall M., Simmons D., Luan X., Nydegger J., Feng J.Q., Gu T.T.;

InterPro: IPR000710; IgA\_S6.

DR InterPro: IPR004899; Pfamact\_sup.

DR Pfam: PF02395; IGAL; 1.

DR Pfam: PF03212; Pertactin; 2.

DR PRINTS: PR00921; IGASERPASE.

KW Hydrolase; Serine protease; Transmembrane; Zymogen; Signal.

FT SIGNAL 1 25 POTENTIAL.

FT CHAIN 26 1021 IMMUNOGLOBULIN A1 PROTEASE.

FT PROPEP 1022 1849 HELPER PEPTIDE (POTENTIAL).

FT ACT\_SITE 299 299 PROBABLE.

SEQ SEQUENCE 1849 AA; 202957 MW; 79A7D018C7150AEA CRC64;

Query Match 3.9%; Score 175.5; DB 1; Length 1849;

Best Local Similarity 21.6%; Pred. No. 0.16;

Matches 106; Conservative 68; Mismatches 218; Indels 99; Gaps 19;

QY 250 NKSCRVT-----TSLCYSDEGQELIVSYSDVIYLFDPKDDTARELKTPSAEERR 299

Db ||||| : | : | : | : | : | : | : | : | : | : | : | : | :

QY 921 NNKSNNVVNVSATGNFTLVADKTGP-----NHNELTLFDSANTRNLLEVLTVANGSV 975

Db :

QY 300 EEIQQPVKKRLRGDNLS----DTGPR---ARPESERERGQSPNVSLMQMSDMLSRW 352

Db :

QY 976 D-----KGAKYKLNRVNGRYDYLNPEVEKRNTVDITNPNDI----- 1016

Db :

QY 353 FEEASEVAQSRRGRSRRGGTSQSDISTLT--PTVPSSPDLEVSTAMEVDTPAEQFLQ 410

Db : : : ||||| :

QY 1017 -QADAPSAQSN-----NEEIARVETVPVPPAPATESAJAQEPETRAETAQAQ 1062

Db :

QY 411 P-----STGSTWSAQAHSTPSHPSTPLSSPDSQRQSVEASGHHTHQSDNNKE 464

Db :

QY 1063 PAMETWTNSTAPKSDATQTGTENPNSE---SVPSETTKVAENPQNETVAKNQEQ 1119

Db :

QY 465 KLSPKPGTG-----PVLHLHYSTEGTTTSTIKLFNDEMSSIASSRSGISGCKSQG 1120

Db :

QY 1120 ATPTONGEVAKEDOPTVEANTQTNATQSGK---TEETQATSKSEPSTVSENQ 1176

Db :

QY 519 EEFVFPQSSQPPEGSEDKAPEESESDVKYQGVSAENVHNINTQSDKFETAK-PL 577

Db :

QY 1177 PERTVSQSTDKVVEKEKAKE-TEETQAPQVTSKEPP-----KQAEPAPEVPT 1228

Db :

QY 578 DNSGGERDLNLRSCGVPESAEEKAKEPETSDQTSSESATNNTNPEPQPTATG 637

Db :

QY 1229 DTNAEAQAQLQOQPTTVAATAETTSPNSKPAAETQAPS-----EKTAPE 1275

Db :

QY 638 PSAHEETSRSALQTDSDSDDPVL-LPGARYAGDGRRSARAVARIQEFRFRKERKEM 696

Db :

QY 1276 PVVSENTATQPTETEETAKYEKETQEVPOVASQESPQBQPA-AKPAQTKPQAEARE 1334

Db :

QY 697 ELDDLNIIRRP 707

Db :

QY 1335 NVLTKNVGEP 1345

Db :



Wed Mar 26 09:38:55 2003

FT MOD\_RES 314 314 PHOSPHORYLATION (BY CK2) (POTENTIAL).  
 FT MOD\_RES 336 336 PHOSPHORYLATION (BY CK2) (POTENTIAL).  
 FT MOD\_RES 349 349 PHOSPHORYLATION (BY CK2) (POTENTIAL).  
 FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 84 84 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 130 130 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 190 190 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 313 313 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 373 373 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 934 934 A618789D8A57249A CRC64;  
 SQ SEQUENCE 934 AA; 93901 MW; 1818789D8A57249A CRC64;  
 Query Match 3.9%; Score 175; DB 1; Length 934;  
 Best Local Similarity 19.6%; Pred. No. 0.079;  
 Matches 84; Conservative 78; Mismatches 205; Indels 62; Gaps 9;  
 QY 262 SEDQELTVSYSDYVLFDPKDDTARELKTSPAEERREELRQPPVKRLRLRGDMSDTGP 321  
 Db 525 SEDKDE---SDSDH---DNSSDS--ESKSDSDSDSDS-----SDSDSDS- 562  
 QY 322 RARPERERDGGQSPNVSLMQRMDSMLSNFEEASVAGSNRGRSRPRGTSQSDIS 381  
 Db 563 ---DSSDSDSDSDSDS-----DSSDSDSDSDSDSDSDSDSDSDSDSDSDS 604  
 QY 382 -----TLPTVPSPDLEYSTAMEYDTTAEQFLOPSTSTMSAQAHSTSPESHST 434  
 Db 605 DTCSDS 664  
 QY 435 PLLSGPDEQSQVETASGHHTHHQSDNNKLSFKPGTGEVPLSLHYSTEGTTSTIKLN 494  
 Db 665 DSSDS 724  
 QY 495 FDEWSS-----IASSRGIGSHCKSEGESEFVQSQVPEGDSKAPESSEDTVK 549  
 Db 725 SDS 784  
 QY 550 YEGVSAENPVENHINITQDKFTAKPLDSNGERNLNLDRSCGVPEASSEKAKEPE 609  
 Db 785 SDS 844  
 QY 610 TSQSTESATNENNTNPFQFQTEATGSAHEETSTRDSALQDTHDDDDVPLPGARY 669  
 Db 845 SDS 895  
 QY 670 RAGPDERS 678  
 Db 896 -SDGDSKS 903  
 RESULT 12  
 ID NASP\_HUMAN STANDARD; PRT; 787 AA.  
 AC P49321; 1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Nuclear autoantigenic sperm protein (NASP).  
 DE NASP.  
 GN Homo sapiens (Human).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC Tissue-Testis;  
 RX MEDLINE=93050782; PubMed=1426632;  
 RA O'Rand M.G., Richardson R.T., Zimmerman L.J., Widgren E.E.;  
 RT "Sequence and localization of human NASP: conservation of a xenopus  
 histone-binding protein."  
 RL Dev. Biol. 154:37-44(1992).  
 CC -1- FUNCTION: MAY PLAY A ROLE IN REGULATING THE EARLY EVENTS OF  
 SPERMATOGENESIS BY BINDING NEWLY SYNTHESIZED HISTONES VARIANTS AND  
 TRANSPORTING THEM TO THE NUCLEUS.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.

CC -1- TISSUE SPECIFICITY: TESTIS- AND SPERM-SPECIFIC PROTEIN.  
 CC -1- SIMILARITY: WITH XENOPUS HISTONE-BINDING PROTEIN N1/N2.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M97856; AAA36027.1; .  
 DR GenBank: HGNC:7644; NASP.  
 DR MIM: 603185;  
 DR InterPro: IPR001440; TPR.  
 DR Pfam: PF00515; TPR; 3.  
 DR SMART: SM00028; TPR; 3.  
 KW Nuclear protein; Sperm; Coiled coil; Antigen.  
 FT DOMAIN 136 164 COILED COIL (POTENTIAL).  
 FT DOMAIN 460 487 COILED COIL (POTENTIAL).  
 FT DOMAIN 597 664 COILED COIL (POTENTIAL).  
 FT DOMAIN 752 769 COILED COIL (POTENTIAL).  
 FT DOMAIN 116 126 GLU-RICH (ACIDIC).  
 FT DOMAIN 482 511 GLU-RICH (ACIDIC).  
 FT DOMAIN 715 721 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT SEQUENCE 787 AA; 85130 MW; 18DBA6DFD48321E0 CRC64;  
 SQ SEQUENCE 787 AA; 85130 MW; 18DBA6DFD48321E0 CRC64;  
 Query Match 3.9%; Score 174.5; DB 1; Length 787;  
 Best Local Similarity 19.5%; Pred. No. 0.068;  
 Matches 126; Conservative 96; Mismatches 228; Indels 195; Gaps 29;  
 QY 283 KDTARELKTSPAEERREELRQPPVKRLRLRGDMSDTGP--ARPERERDEQ----- 335  
 Db 124 EDESLVNNNDIDEAREELREQVYDAMGEKAEKTKSLAKPETDKEDSEMEKGR 183  
 QY 336 -----SPNVSLMQRMDSMLSNFEEASVAGSNRGRSRPRGTSQSDISTLTTPVSSP 390  
 Db 184 EMDISKSAEPEQEKVDTLDTLDTETSEE-----KGAAPG-EPNEAETSGKPEQEV 237  
 QY 391 DLE-----VSTAMEYDTTAEQFLOPSTSTMSAQAHSTSPST 428  
 Db 238 DAEEKSVSGTDVQECREKGGQKQGEVIVSIEKPEVSEQPPVTLKQGTAVEVA 297  
 QY 429 ES--PHSTPLSSPDSEQRQSVASGHHHTHQSDNNKLSKPKGTGEPVLSLHYSTEGT 486  
 Db 298 ESLDPTVRPVDVGDEPEKVV-----TSENAKAVLEQLVGOEVPFAEPEVQ 348  
 QY 487 TTSTIKLNTDEWSSIASSSRGIGSHCKSEGESEFVQ--SSVQPP--GDSEKAPES 543  
 Db 349 TEAA-----EASAVEAGS-----EVSEKPGQEAFLPGAVNGPSVVGDTPIEPQTS 397  
 QY 544 SEDVTYKQGVSAENPVENHINITQDKFTAKPLDSNGERNLNLDRSCGVPEASSE 603  
 Db 398 IERLTETKDGSGLEEKV-----RAKLVPSQEEATKL 427  
 QY 604 KAKEPETS--DOTSSTATNENNTNPFQFQTEATGSAHEETSTRDSAL---QDTDDSD 659  
 Db 428 SVESEAGDGVDTKVAQGAATEKSPEDKVQI-----AAEETQEREQKMEETEGSEE 482  
 QY 660 DPLVPGARYRAGPDERSAVARIQEFFRRRKERKEMEELDTLNI---RRLPVKMYKGH 716  
 Db 483 DD-----KENDTEEMPNDVSLNKSLOENEEEEEIGLEAWDMLDLAKIIFK-- 530  
 QY 717 RNSRTMIKEANFWANFVNSGDCGHIFWDRHTAEHML-----LEADNHV-----VNC 766  
 Db 531 ---RQETKEAQLYAAQ-----AHLKLGVSVESENYQVAVEEFCQ 568  
 QY 767 -----LQPPFPFILASS-----GIDYDIKWSPLESRIFNKKLADEVITRNLML 815  
 Db 569 LNLQEQYLEAH--DRLLAETHYQLGLAYG-----YNSQY-DEAVAQFSKIEV 613  
 QY 816 TRNTITVPASFLMLRLASLNHRADRLBEGDSESGSQENENEDEE 860



DB 614 IENRMAV-----LNE-----QVKEAGS-SEYKKEIEE 640

RESULT 13

MOZ\_HUMAN

AC Q92794; STANDARD; PRT; 2004 AA.

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Monocytic leukemia zinc finger protein (zinc finger protein 220).

GN ZNF220 OR MOZ.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=96376968; PubMed=8782817;

RA Borrow J., Stanton V.P. Jr., Andresen J.M., Becher R., Behm F.G.,

RA Chaganti R.S.K., Civin C.I., Distche C., Dube I., Frischauf A.M.,

RA Horman D., Mitelman F., Volinia S., Wamere A.E., Housman D.E.,

RT "The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses

RT a putative acetyltransferase to the CREB-binding protein.";

RL Nat. Genet. 14:33-41(1996).

CC -!- FUNCTION: MAY REPRESENT A CHROMATIN-ASSOCIATED ACETYLTRANSFERASE.

CC -!- SUBCELLULAR LOCATION: Nuclear.

CC -!- DISEASE: PARTICIPATES IN A T(8;16)(P11;P13) CHROMOSOMAL

CC TRANSLOCATION THAT PRODUCES A MOZ-CBP CHIMAERA OBSERVED IN THE

CC M4/M5 SUBTYPE OF ACUTE MYELOID LEUKEMIA.

CC -!- SIMILARITY: CONTAINS 2 PHD-TYPE ZINC FINGERS.

CC -!- SIMILARITY: BELONGS TO THE MIST (SAS/MOZ) FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; U47742; AAC50662.1;

DR Genbank; HGNC:13013; ZNF220.

DR MIM; 601408;

DR InterPro; IPR001386; Histone\_H1/H5.

DR InterPro; IPR002717; MOZ\_SAS.

DR InterPro; IPR001965; Znf\_PHD.

DR Pfam; PF00628; PHD; 2.

DR Pfam; PF01853; MOZ\_SAS; 1.

DR SMART; SM00526; H15; 1.

DR SMART; SM00249; PHD; 2.

DR PROSITE; PS01359; ZF\_PHD\_1; 1.

DR PROSITE; PS50016; ZF\_PHD\_2; 2.

KW Proto-oncogene; Chromosomal translocation; Zinc-finger; Repeat;

KW Nuclear protein.

FT ZN\_FING 206 265 PHD-TYPE 1.

FT ZN\_FING 259 313 PHD-TYPE 2.

FT DOMAIN 371 379 POLY-SER.

FT ZN\_FING 538 560 C2HC-TYPE.

FT DOMAIN 788 801 POLY-GLU.

FT DOMAIN 989 995 POLY-GLU.

FT DOMAIN 1019 1026 POLY-ARG.

FT DOMAIN 1069 1078 POLY-GLU.

FT DOMAIN 1147 1150 POLY-LYS.

FT DOMAIN 1221 1242 GLU-RICH.

FT DOMAIN 1267 1302 GLU-RICH.

FT DOMAIN 1411 1414 POLY-GLU.

FT DOMAIN 1593 1597 POLY-SER.

FT DOMAIN 1643 1704 GLN/PRO-RICH.

FT DOMAIN 1897 1977 MET-RICH.

FT SITE 1546 1547 BREAKPOINT FOR TRANSLOCATION TO FORM

SEQUENCE 2004 AA; 225054 MW; 9FTBBAC3792854BA CRC64;

Query Match 3.9%; Score 174.5; DB 1; Length 2004;

Best Local Similarity 19.7%; Pred. No. 0.2;

Matches 115; Conservative 84; Mismatches 231; Indels 155; Gaps 23;

QY 261 YSEDGOEILVSVSSDIYILFDPKDDTARELKTSPASEREREELRQPPVKRLRLR-GDWST 319

DB 913 YTESEQLVASEEQP---SQDGKPLPKRLSEGVPWGQLKSP-EALKRRLTEGSR 968

QY 320 GPARPESR-----ERDGEOSPVS-----LMORMDMLSRWFEEASEVA 360

DB 969 LFRRYSEGDAVLRGFSESSESESPRSSPPILKTKRKKPFLHR----- 1019

QY 361 QSNRGGRSRRGTS-----QSDISTLPTVPSSPDLEYSETAMEVDT 403

DB 1020 ---RRVRKRKHNSVVTETISETTEVLDPEDSDSERPMPLPTEFEDEEEDE 1076

QY 404 ---PAEQFLQPTSTSTMSQAHS-----TSPSTESPHSTPLSSPDSEQRSVEAS 451

DB 1077 NELFPREYFRLLSQDVLRCOSSSKRKSKDEEEDSDADDTPILKPVSLLRKRDVKS 1136

QY 452 GHHTHGHSDNNNEKLSKPGTGEPVLSLHYSTEGTTTSTIKLNTDFWSSIASRGI-- 509

DB 1137 PLEPDTSTPLKKKGWPKGSKRKPI---HWKRPGRKPGFKL--SREIMPVSTQACVIEP 1191

QY 510 -----GSHCKSEGOSESFVPOSSVOPPEGSDSETKAPESSEDVTKYQEG-----VSAE 557

DB 1192 IVSIPKAGRKPKIQESEETVEPKEDMPLP-----ERKEEEMQAEAEAEDEEADAAS 1248

QY 558 NPVENHINITQSKFKAKPLDSNSGRNLDNRSCGVPSEASSEKAKEPETSOTSTE 617

DB 1249 VP-----AASPADSNS-----PE-----TETKEPEVEEEKEP 1277

QY 618 SATNENNTNPEQFQTEATGPSAHEETSPDSALQDTSDDDPVLIPGAR----- 668

DB 1278 RYSEQRQSEEEQLEPEPEPEEEDAAETQNDHDDADDEDDHLESTKKELEEQPT 1337

QY 669 ---YRAGPDRRSVARTQEFRRKEREKMEELDLNIRPLVWYKVGHRNSMTIKE 725

DB 1338 REDVKEEFGVQSFADANQMSREKIKDELTDS-BEEQP-----SHTSVVSE 1387

QY 726 ANFWGANFVMSGDCGHFIWDRHTAEHLMLLEADNHVNCLOPH 770

DB 1388 Q-----MAGSEDDH--EEDSHTKEELIELKEEEI-----PH 1417

RESULT 14

ATRX\_HUMAN

ID ATRX\_HUMAN STANDARD; PRT; 2492 AA.

AC P46100; P51068; Q15886; Q9NTS3; Q9H0Z1;

DT 01-NOV-1995 (Rel. 32, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DE Transcriptional regulator ATRX (X-linked helicase II) (X-linked

DE nuclear protein) (XNP) (Znf-HX).

GN ATRX OR RAD54L OR XH2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5), VARIANT S-1860, AND

RP VARIANTS ATR-X.

RX MEDLINE=97123494; PubMed=8968741;

RA Picketts D.J., Higgs D.R., Bachoo S., Blake D.J., Quarrell O.W.J.,

RA Gibbons R.J.;

RT "ATRX encodes a novel member of the SNF2 family of proteins: mutations

RT point to a common mechanism underlying the ATR-X syndrome.";

RL Hum. Mol. Genet. 5:1899-1907(1996).

CC [2]

RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 4).

RX MEDLINE=97386582; PubMed=9244431;

RX Villard L., Lossi A.-M., Cardoso C., Proud V., Chiaroni P.,

Wed Mar 26 09:38:55 2003

- RA Colleaux L., Schwartz C., Fontes M.;  
 RT "Determination of the genomic structure of the XNP/ATRX gene encoding  
 RL a potential zinc finger helixase.";  
 RN Genomics 43:149-155(1997).  
 [3]  
 RN SEQUENCE OF 860-2492 FROM N.A.  
 RP MEDLINE-95179111; PubMed-7874112;  
 RX Stayton C.L., Dabovic B., Gullisano M., Gecz J., Broccoli V.,  
 RA Stayton C.L., Dabovic B., Gullisano M., Monaco L., Rastan S., Boncinelli E.,  
 RA Giovanazzi S., Bossolasco M., Gecz J., Broccoli V., Boncinelli E.,  
 RA Bianchi M.E., Consalez G.G.;  
 RT "Cloning and characterization of a new human Xql3 gene, encoding a  
 RL putative helixase.";  
 RN Hum. Mol. Genet. 3:1957-1964(1994).  
 [4]  
 RN PRELIMINARY PARTIAL SEQUENCE FROM N.A.  
 RP MEDLINE-94214473; PubMed-8162050;  
 RX Gecz J., Pollard H., Consalez G., Villard L., Stayton C.L.,  
 RA Milasseau P., Khrestchatsky M., Fontes M.;  
 RT "Cloning and expression of the murine homologue of a putative human  
 RL X-linked nuclear protein gene closely linked to PGK1 in Xql3.3.";  
 RN Hum. Mol. Genet. 3:39-44(1994).  
 [5]  
 RN SEQUENCE OF 2401-2492 FROM N.A., AND VARIANTS ATR-X.  
 RP MEDLINE-95211835; PubMed-7697714;  
 RX Gibbons R.J., Picketts D.J., Villard L., Higgs D.R.;  
 RA "Mutations in a putative global transcriptional regulator cause X-  
 RT linked mental retardation with alpha-thalassemia (ATR-X syndrome).";  
 RL Cell 80:837-845(1995).  
 [6]  
 RN SEQUENCE OF 1375-2492 FROM N.A.  
 RP Pearce A., Chapman J.;  
 RA Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 [7]  
 RN EZH2 BINDING.  
 RP MEDLINE-96167853; PubMed-9499421;  
 RX Cardoso C., Timsit S., Villard L., Khrestchatsky M., Fontes M.,  
 RA Colleaux L.;  
 RT "Specific interaction between the XNP/ATRX gene product and the SET  
 RL domain of the human EZH2 protein.";  
 RN Hum. Mol. Genet. 7:679-684(1998).  
 [8]  
 RN SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC  
 RP HETEROCHROMATIN.  
 RX MEDLINE-20406663; PubMed-10570185;  
 RA McDowell T.L., Gibbons R.J., Sutherland H., O'Rourke D.M.,  
 RA Buckmore W.A., Pombor A., Turley H., Gatter K., Picketts D.J.,  
 RA Buckle V.J., Chapman L., Rhodes D., Higgs D.R.;  
 RT "Localization of a putative transcriptional regulator (ATRX) at  
 RL pericentromeric heterochromatin and the short arms of acrocentric  
 RN chromosomes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:13983-13988(1999).  
 [9]  
 RN DISEASE.  
 RP MEDLINE-20213147; PubMed-10751095; Gecz J.;  
 RX Villard L., Fontes M., Ades L.C.;  
 RA "Identification of a mutation in the XNP/ATRX gene in a family  
 RT reported as Smith-Fineman-Myers syndrome.";  
 RL Am. J. Med. Genet. 91:83-85(2000).  
 [10]  
 RN VARIANT ATR-X SER-1713.  
 RP MEDLINE-97196774; PubMed-9043863;  
 RX Villard L., Lacombe D., Fontes M.;  
 RA "A point mutation in the XNP gene, associated with an ATR-X phenotype  
 RT without alpha-thalassemia.";  
 RL Eur. J. Hum. Genet. 4:316-320(1996).  
 [11]  
 RN VARIANT JM GLN-2131.  
 RP MEDLINE-96224392; PubMed-8630485;  
 RX Villard L., Gecz J., Mattei J.-F., Fontes M., Saugter-Weber P.,  
 RA Munnich A., Lyonnet S.;  
 RT "XNP mutation in a large family with Juberg-Marsidi syndrome.";  
 RL Nat. Genet. 12:359-360(1996).  
 [12]  
 RN VARIANTS ATR-X.  
 RP MEDLINE-97467722; PubMed-93269311;  
 RX Gibbons R.J., Bachoo S., Picketts D.J., Aftimos S., Aisenbauer B.,  
 RA Bergoffen J., Berry S.A., Dahl N., Fryer A., Keppler K., Kurosawa K.,  
 RA Levin M.L., Masuno M., Neri G., Pierpont M.E., Slaney S.F.,  
 RA Higgs D.R.;  
 RT "Mutations in transcriptional regulator ATRX establish the functional  
 RL significance of a PHD-like domain.";  
 RN Nat. Genet. 17:146-148(1997).  
 [13]  
 RN VARIANT ATR-X LEU-246.  
 RP MEDLINE-20123062; PubMed-10660327;  
 RX Fichera M., Romano C., Castiglia L., Pailla P., Ruberto C., Amata S.,  
 RA Greco D., Cardoso C., Fontes M., Ragusa A.;  
 RA "New mutations in XNP/ATRX gene: a further contribution to  
 RT genotype/phenotype relationship in ATR/X syndrome.";  
 RL Hum. Mutat. 12:214-214(1998).  
 [14]  
 RN VARIANT SHS LYS-1742.  
 RP MEDLINE-99347960; PubMed-10417298;  
 RX Lissi A.-M., Millan J.M., Villard L., Orellana C., Cardoso C.,  
 RA Prieto F., Fontes M., Martinez F.;  
 RA "Mutation of the XNP/ATRX gene in a family with severe mental  
 RT retardation, spastic paraplegia and skewed pattern of X inactivation:  
 RL demonstration that the mutation is involved in the inactivation  
 RN bias.";  
 RL Am. J. Hum. Genet. 65:558-562(1999).  
 [15]  
 RN VARIANT CWS THR-2050.  
 RP MEDLINE-99326061; PubMed-10398237;  
 RX Abidi F., Schwartz C.E., Carpenter N.J., Villard L., Fontes M.,  
 RA Curtis M.;  
 RA "Carpenter-Waziri syndrome results from a mutation in XNP.";  
 RL Am. J. Med. Genet. 85:249-251(1999).  
 [16]  
 RN VARIANTS ATR-X E-175; 178-V-K-198 DEL; S-190; P-219; L-246 AND C-249.  
 RP MEDLINE-99219535; PubMed-10204841;  
 RX Villard L., Bonino M.-C., Abidi F., Ragusa A., Belouigne J.,  
 RA Lissi A.-M., Seaver L., Nefont J.-P., Romano C., Fichera M.,  
 RA Lacombe D., Hanauer A., Philip N., Schwartz C.E., Fontes M.;  
 RT "Evaluation of a mutation screening strategy for sporadic cases of  
 RL ATR-X syndrome.";  
 RL J. Med. Genet. 36:183-186(1999).  
 [17]  
 RN VARIANTS ATR-X S-179; L-190; C-246; F-1552; S-1645 AND C-1847.  
 RP MEDLINE-20451413; PubMed-10995512;  
 RX Wada T., Kubota T., Fukushima Y., Saitoh S.;  
 RA "Molecular genetic study of Japanese patients with X-linked  
 RT alpha-thalassemia/mental retardation syndrome (ATR-X).";  
 RL Am. J. Med. Genet. 94:242-248(2000).  
 CC -1- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES  
 CC GENE EXPRESSION BY AFFECTING CHROMATIN. MAY BE INVOLVED IN  
 CC BRAIN DEVELOPMENT AND FACIAL MORPHOGENESIS.  
 CC -1- SUBUNIT: PROBABLY BINDS EZH2. BINDS ANNEXIN V IN A CALCIUM AND  
 CC PHOSPHATIDYLCHOLINE/PHOSPHATIDYLSELINE-DEPENDENT MANNER (BY  
 CC similarity).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR. ASSOCIATED WITH PERICENTROMERIC  
 CC HETEROCHROMATIN DURING INTERPHASE AND MITOSIS, PROBABLY BY  
 CC INTERACTING WITH HP1.  
 CC -1- ALTERNATIVE PRODUCTS: 5 ISOFORMS; 1, 2, 3, 4 (SHOWN HERE) AND 5;  
 CC ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: UBICUITOUS.  
 CC -1- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF X-LINKED ALPHA-  
 CC THALASSEMIA/MENTAL RETARDATION SYNDROME (ALSO KNOWN AS ATR-X  
 CC SYNDROME). ATR-X IS AN X-LINKED DISORDER COMPRISING SEVERE  
 CC PSYCHOMOTOR RETARDATION, FACIAL DYSMORPHISM, UROGENITAL  
 CC ABNORMALITIES, AND ALPHA-THALASSEMIA. AN ESSENTIAL PHENOTYPIC  
 CC TRAIT ARE HEMOGLOBIN H ERYTHROCYTE INCLUSIONS.  
 CC -1- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF CARPENTER-WAZIRI  
 CC SYNDROME (CWS), AN X-LINKED RECESSIVE CONDITION CHARACTERIZED BY  
 CC MODERATE MENTAL RETARDATION, SHORT STATURE, BRACHYDACTYLY WITH  
 CC EXCESSIVE SKIN CREASES, AND WIDENING OF THE KNUCKLES.  
 CC -1- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF JUBERG-MARSIDI SYNDROME

(JM). JM IS A RARE X-LINKED RECESSIVE DISEASE CHARACTERIZED BY SEVERE MENTAL RETARDATION, GROWTH FAILURE, SENSORINEURAL DEAFNESS, MICROGENTALISM AND EARLY DEATH.

-1- DISEASE: DEFECTS IN ATRX ARE A CAUSE OF SMITH-FINEMAN-MYERS SYNDROME (SFM). CLINICAL FEATURES INCLUDE SEVERE MENTAL RETARDATION, MICROCEPHALY, GROWTH FAILURE, FACIAL ANOMALIES AND BILATERAL CRYPTORCHIDISM, DUE TO THE CLINICAL OVERLAP WITH ATR-X SYNDROME, SOME PATIENTS ORIGINALLY DIAGNOSED AS HAVING SFM, MIGHT BE AFFECTED BY A VARIANT OF ATR-X SYNDROME WHICH LACK HEMOGLOBIN H INCLUSIONS.

-1- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF SUTHERLAND-HAAN X-LINKED MENTAL RETARDATION SYNDROME (SHS). IT IS CHARACTERIZED BY SEVERE STATURE AND CRYPTORCHIDISM.

-1- SIMILARITY: BELONGS TO THE SNE2/RAD54 HELICASE FAMILY.

-1- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.

-----

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; U72937; AAB49970.2; ;  
 DR EMBL; U72938; AAB49971.2; ;  
 DR EMBL; U72935; AAB40698.1; ;  
 DR EMBL; U72904; AAB40698.1; JOINED.

Query Match  
 Best Local Similarity 3.8%; Score 173; DB 1; Length 2492;  
 Matches 125; Conservative 93; Mismatches 226; Indels 232; Gaps 26;

QY 248 HLNNKCRVTLSCYSEGGQILVSYSDYILFDPKDDTARELKTSAERREELRQPPV 307  
 Db 948 HLTKTKKKVQDGLSDIAEFLKQSD-----ETSEDDKKQSKKTEKKPKSPFKKV 1002  
 QY 308 KRLRLGWSDTGPRAPRESER-----ERDGGSPNVLORMSDMLSRWF 353  
 Db 1003 IKMEQYESSDGTLEPLREITCHPPGKIKQKNGTTDGEKK-----SKLRDKTSKK 1057  
 QY 354 EEASEVAQNRG-----RSRPGGTQSD---TST 382  
 Db 1058 DELSDYAEKSTGKSDGSSDDKSKNGAYGREKKRCLLGGSSKRDGSSSTERYSM 1117  
 QY 383 LPTVPSPDLVSETPAMEVDTPAEQFLQPSSTMSAQAHSTSPSTPHSTPLSSPDS 442  
 Db 1118 KEDGCSNDDKRLKRIELR-----ERNLSSKRNTKEIQSGSSSDAEE-----SSEDN 1165  
 QY 443 EORQSVASGHHT--HQSDNNNEKLPKPGTGEPVLISLHYSTGTTTSTIKLNFDTBW- 499  
 Db 1166 KKKQRTSSKKKAVIVKEKKRNLSTTKRKQADITSSSSDIEDDDQNSIGGSSDEQK 1225  
 QY 500 -----SIASSSRGIGSHCKSEGE--ESFVP-----QSS 526  
 Db 1226 IKPVTENLVLSSTGTF---CQSSDEALSKSPVTVDDDDDDNDPENRIAKMLLEIKR 1282  
 QY 527 SVQPPGDSSTKAPESSSDVTKQGVGAENPVENHINITQSD---KFTAKP-----576  
 Db 1283 NLSSDEGSSDDPEGKKTGKQNEENFNGDEAKQNVN-SESDSDSESKKPRYHRL 1341  
 QY 577 -----LDNSNGE-----RN-----DLNDRSCGVPESSASEKAK 606  
 Db 1342 RHLTVSDGESGEKTKPKHEKVRNRKVVSSDESDSDSFQES-GVSEVSEDEQ 1400  
 QY 607 EPET-----SDQTS-----TESATNENNTNPEQFOFQTEATGPSAHEITS 645  
 Db 1401 RPRTRSAAKAELEENQRYKQKKRRRIKVQDSSSENKSNSEEEEEK-----EEEE 1454  
 QY 646 TRDSALQDTSDDDPVLIPGARYRAGPGDRRSVART-----Q 684  
 Db 1455 EEEEEEEEDEND-----SKSPGKRRKIRKILDKDLRTETQNALKEEERRK 1505

QY 685 EFFRRRKRERKEEL-----DTLNIRRLPLVKMYKGHNSRTMIK 724  
 Db 1506 RIAERERERKLEVEIEDASPTKPIITTKLVLDDEETKEPLQV-----HRNMVIRKLK 1561  
 QY 725 EANFWGANFVMSGSDC 740  
 Db 1562 PHQVGVQVFMW---DC 1574

RESULT 15  
 DMPI\_BOVIN  
 ID DMPI\_BOVIN STANDARD; PRT; 510 AA.  
 AC Q95120;  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Dentin matrix acidic phosphoprotein 1 precursor (Dentin matrix protein-1) (DMP-1).  
 GN DMPI.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Tooth;  
 RX MEDLINE=97263952; PubMed=9109824;  
 RA Hirsh K.L., Ibaraki-O'Connor K., Young M.F., Dixon M.J.;  
 RT "Cloning and expression analysis of the bovine dentin matrix acidic phosphoprotein gene.";  
 RL J. Dent. Res. 76:754-760(1997).  
 CC -1- FUNCTION: COULD BE INVOLVED IN THE INDUCTION OF MINERALIZATION OF EXTRACELLULAR MATRIX AND IN DENTINOGENESIS.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN FETAL BRAIN, BONE AND TOOTH PARTICULARLY IN ODONTOBLAST, BUT NOT IN AMELOBLAST. NOT EXPRESSED IN LIVER AND SKIN.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; U47636; AAB09412.1; ;  
 DR Extracellular matrix; Signal.  
 KW SIGNAL 1 16 POTENTIAL.  
 FT CHAIN 17 510 DENTIN MATRIX ACIDIC PHOSPHOPROTEIN 1.  
 FT CARBOHYD 351 351 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 370 370 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 427 427 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 464 464 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT SITE 364 366 CELL ATTACHMENT SITE (POTENTIAL).  
 SQ SEQUENCE 510 AA; 55491 MW; 9BFA9A74F6450865 CRC64;

Query Match  
 Best Local Similarity 3.8%; Score 172; DB 1; Length 510;  
 Matches 102; Conservative 49; Mismatches 193; Indels 90; Gaps 18;

QY 281 DPKDDTARELTPSAERREELRQPPVKRLRLG-----DMSDT 319  
 Db 65 DPSDSTSE-EVLGLDQOQVHR--PAGGLSRGGSGDKDDDESGDDTDFGDDGGP 121  
 QY 320 GPRAPRESERDEGEQSPNVLORMSDMLSRWFEEASEVAQSNR-----GRGRSRPRGG 374  
 Db 122 GPERRRSGGSRGSDSDSADTTTSREDSTPQGGEGARTTSESRLDRDEGNSRPEG 181  
 QY 375 TSQSDISTLTP-----TVPSSPDLEVSETAMEYDVTAEQFLQPSSTMSAQAHSTS 425  
 Db 182 DSTPDSSEHWVGSGSGDSSHGGSEFFDEGCMQSDDPGATRSERGNRISDAGLKTQ 241

DR EMBL; U89012; AAC51332.1; -;  
DR EMBL; U34037; AAB97602.1; -;  
DR EMBL; U65378; AAB87728.1; -;  
DR Genew; HGNC:2932; DMPL.  
DR MIM: 600980; -;  
KW Extracellular matrix; Signal; Alternative splicing.  
FT SIGNAL 1 16 POTENTIAL.  
FT CHAIN 17 513 DENTIN MATRIX ACIDIC PHOSPHOPROTEIN 1.  
FT CARBOHYD 25 25 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 285 285 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 324 324 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 345 345 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 351 351 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 413 413 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 426 426 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 467 467 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT SITE 364 366 CELL ATTACHMENT SITE (POTENTIAL).  
FT VARSPLIC 46 61 MISSING (IN ISOFORM 2).  
FT CONFLICT 69 69 S -> C (IN REF. 2).  
SQ SEQUENCE 513 AA; 55782 MW; 2C1FDE319A5D106F CRC64;  
Query Match 3.8%; Score 171.5; DB 1; Length 513;  
Best Local Similarity 21.8%; Pred. No. 0.059;  
Matches 102; Conservative 58; Mismatches 210; Indels 97; Gaps 18;  
QY 262 SEDGQELLVSYSDVYILF-----DPKDDTARELKTPSAEERREELRQPPVKR 309  
Db 69 STQSEGLGSDDDHQIYIRLAGFSRSTGKGDDKDD---EDDGGD----- 112  
QY 310 LRLRGWSDTGPRAPESERERDGEQSPNVSLMQRMDSMLRWFEEASEVAQ-----SNR 364  
Db 113 -TFGDDSGGPKDRQEGGNSRUGSDESDDTIQAEESAPQGOQAQDTTSESRELDNE 171  
QY 365 GRGRSRPRGCTS-QSDISTLTFTVPSPD-----LEVSETAMEVDTPAEQFLQPSSTSS 415  
QY 172 DRVDSKPEGGDSQSESEEHVWGGGSDGESSHGDSGLDDEGCMQSDDP-ESIRSERGNS 230  
Db 416 TMSAQ-----HSTSSPTEGPHSPPLSSPD-----SEORQSVASGHHTHQ 458  
QY 231 RMNSAGMKSKESGENSEQANTQDSGGGQLLEHPSRKIFTKRSRISEDDRSELDDNNITMEE 290  
QY 459 --SDNNNEKLSPKPGTGPVLSLHYSTGTTTSTIKLFTDEW-----SSIASSSRGI 509  
Db 291 VKSDSTENSNSRDTGLSQP----RDSKGDQSDQSKENLQSEESQNVDPSPSESSQEANL 346  
QY 510 GHCKSEQEQEESFVQSSVQPP-----EGDSETKAPESSEDDVTYQEGVSANPV-- 560  
Db 347 SSQENSSSEQEEVSESRGNDPPTTSYVEDQSDSSEDSSTLSHKSSESREQADS 406  
QY 561 ENHINTQDQKFTAKPLDSNGERNDLNRCSGPVEASAKAEKPESTD-QTSTESA 619  
Db 407 ESSESLNFSESPESPEDNSSQEGLOSHSSA---ESQSESHSEEDSDSDSQDSRSRK 463  
QY 620 TNENNTNPFOFTEATGPSAEETSTR-----DSALQDTDSD 658  
Db 464 EDNST--ESKSSSEEDGQLKIEIESRKLTVDAYHNKPIGDDND 508  
RESULT 17  
DXT STDIO  
ID DEXT STDIO STANDARD; PRT: 1337 AA.  
AC P39653; 1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 01-FEB-1995 (Rel. 31, Last annotation update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Dextranase precursor (EC 3.2.1.11) (Alpha-1,6-glucan-6-  
DE glucanohydrolase).  
GN DEX.  
OS Streptococcus downei (Streptococcus sobrinus).  
OC Plasmid pYA902.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1317;

cc or send an email to [licensing@bca.ca](mailto:licensing@bca.ca);  
cc  
cc



Rt laevis are expressed during

529 AEW RVVGQENWHARYDAK-----EVSREGRJVVGL

529 AEWRRVVGQEKWHARYYDAK-----EVSABSLIIVGGR

[illegible]







QY 511 SHCKSEGO-----ESFVPOSSVQPPED-----SETKAPESSEDVTKYQGVSAENPVEN 562  
 Db 1258 KEFVTEKNPVVEE---PSKQEDKPKVEKIAEEOAPVEPEDEDEDESTSTSTTET 1314  
 QY 563 HINIQTOKFTAKP-----LDSNGERNDLNRSCGVPESASS-----EKAKE 607  
 Db 1315 P---SPSGSYTKPGLSLGSRVDEQSGSNKLDYDDNNQWILANRGYDYLTKSHE 1371  
 QY 608 PETSDOTSTESATNEN-----NTNPEQPOF-----TEATGP-----SAHEETSPTDSA 650  
 Db 1372 RQFAGQONLEMPKNCNYDNGSQEQYGPYEPQADEPSTTEGKAKRALSVKQQAQLLNA 1431  
 QY 651 LQDT-DDSD-----DVLTPGARYRAGDGRSARVARI-----QEFERRKKEKEME 697  
 Db 1432 LNDGSDSDGTTTTPPYAMRGKFRVRRSTARRVPIPKIGKASDEVWVRSPQAKMP 1491  
 QY 698 ELDTLNR 707  
 Db 1492 QRPKSKMSRP 1501

RESULT 22  
 ID IE63\_MCMVS STANDARD; PRT; 841 AA.  
 AC Q69154;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DE Transcriptional regulator IE63 homolog (Protein UL69).  
 GN UL69.  
 OS Murine cytomegalovirus (strain Smith).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Betaherpesvirinae; Murine cytomegalovirus.  
 OX NCBI\_TaxID=10367;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96082764; PubMed=7483291;  
 RA Messerle M., Rapp M., Lucin P., Koszowski U.H.;  
 RT "Characterization of a conserved gene block in the murine  
 RT cytomegalovirus genome."  
 RL Virus Genes 10:73-80(1995).  
 CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL54,  
 CC HSV-2 UL54, EBV-1 5, VZV 4, EBV BMLF1, HVS-1 57, HCMV AND MCMV  
 CC UL69.

QY 190 CRRATSVAC-----PPI-----PYLAVGSSSVRIYD-----RRLM----- 224  
 Db 326 CRKGGSSAAAASFDVLRPPFLSKPILQALALIASFARIVGVRRLRHSGFFIND 385  
 QY 225 --GTRATGNAGRTGCMVARTIPSH-LNNKSCRVTSLCYSEGGQELVSYSSDYILFD 281  
 Db 386 FDDTGATDSYRCMISLIDFLDPLGHCRCNEICRVKL-----KLLQPYTS--TLFFC 436  
 QY 282 PKDQTAPE---LKTPSAEERRELPPVKKRLRLGSDSDGTPRARP-----ESER- 329  
 Db 437 AYNTKRHPLNGLPARSPERRAPDTPNIPRLAYRRS-ATTSPEVERAPPSPRMTSSSPRV 495

Query Match  
 Best Local Similarity 3.7%; Score 168; DB 1; Length 841;  
 Matches 132; Conservative 61; Mismatches 205; Indels 164; Gaps 30;

QY 190 CRRATSVAC-----PPI-----PYLAVGSSSVRIYD-----RRLM----- 224  
 Db 326 CRKGGSSAAAASFDVLRPPFLSKPILQALALIASFARIVGVRRLRHSGFFIND 385  
 QY 225 --GTRATGNAGRTGCMVARTIPSH-LNNKSCRVTSLCYSEGGQELVSYSSDYILFD 281  
 Db 386 FDDTGATDSYRCMISLIDFLDPLGHCRCNEICRVKL-----KLLQPYTS--TLFFC 436  
 QY 282 PKDQTAPE---LKTPSAEERRELPPVKKRLRLGSDSDGTPRARP-----ESER- 329  
 Db 437 AYNTKRHPLNGLPARSPERRAPDTPNIPRLAYRRS-ATTSPEVERAPPSPRMTSSSPRV 495

QY 330 -----ERDQSGSPNVLMOR-----MSDMLSRWFEEASEVAQSNRG--RG 367  
 Db 496 DSRGGGDRRGGSSSTSNHHHHTRRARTSTHDSSSSGSRSSRSATDGRRRGSRRG 555  
 QY 368 RSRPGGTQSQSDISTLTPVPSSPDLEYSETAMEVDTPAEQFLOPSTSSSTMSAQAHSTSSP 427  
 Db 556 EAQRESNGHSHSKS-----PST-----VSSITVHGQCARGDSAPSRKSSQSQOQPETTCK 606  
 QY 428 TES-----PHSTPLSSSPDSEORSEASGHHTHQSDNNNEKLSPK---PGTGPV-L 477  
 Db 607 ESSKTRAMPPPPPCSPSPASRRR-----PSKSPSSSPRPHDPSPGEPADA 653  
 QY 478 SLHYSTEG-----TTTSTIKLNTDEWSSSTASSRGISGCKSEGOERSFVPS 526  
 Db 654 EKELATAGDEGVSPGECVATRRGSADESSSSSSSSSSSS-----SSDEESDVEDC 710  
 QY 527 SVQPPGDSSETKAPESSEDVTKYQGVSAENPVENHINITQSKDTAKPLDLSNGSRND 586  
 Db 711 R-----ELDLQSKRLEALEE-----RCERDFE-----ADDEFAEPI-----EEDD 747  
 QY 587 LNLDRSCGVPEESASSEKAKEPETSDQTSATNENNTNPEQFQTEATGSA-----HE 642  
 Db 748 LH-----CSLMEEDIEDPLDPET-ESVWTASVT-----PLAAPPSSIRILDHE 790  
 QY 643 ETSTRDSALQDTD--DSDDDPV 662  
 Db 791 PGDAEEESDSTDYDQPL 812

RESULT 23  
 ID DNJM\_MYCPN STANDARD; PRT; 910 AA.  
 AC P75354;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DE DnaJ-like protein MG200 homolog.  
 GN MPN119 OR MP035.  
 OS Mycoplasma pneumoniae.  
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
 OX NCBI\_TaxID=2104;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 29342 / M129;  
 RX MEDLINE=97105885; PubMed=8948633;  
 RA Himmelfreid R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,  
 RA Hermann R.;  
 RT "Complete sequence analysis of the genome of the bacterium Mycoplasma  
 RT pneumoniae."  
 RL Nucleic Acids Res. 24:4420-4449(1996).  
 CC -1- SIMILARITY: CONTAINS 1 J DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; AE000004; AAB95683.1;  
 DR HSP; P25685; 1HDJ.  
 DR InterPro; IPR001623; DnaJ\_N.  
 DR Pfam; PF00226; DnaJ\_1  
 DR SMART; SM00271; DnaJ\_1  
 DR PROSITE; PS00636; DnaJ\_1;  
 DR PROSITE; PS0076; DnaJ\_2; 1.  
 KW Hypothetical protein; Chapterone; Complete proteome.  
 FT DOMAIN 4 73 J-DOMAIN  
 SQ SEQUENCE 910 AA; 100190 MW; 1250E37D2D221A7 CRC64;

Query Match

3.7%; Score 168; DB 1; Length 910;

us-09-781-693a-2.rsp

Wed Mar 26 09:38:55 2003

```

J. Neurochem. 67:89-97(1996).
-1- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES
    GENE EXPRESSION BY AFFECTING CHROMATIN.
-1- SUBUNIT: PROBABLY BINDS EZH2. BINDS ANNEXIN V IN A CALCULIN AND
    PHOSPHATIDYLCHOLINE/PHOSPHATIDYLSEKINE-DEPENDENT MANNER.
-1- SUBCELLULAR LOCATION: NUCLEAR. ASSOCIATED WITH PERICENTROMERIC
    HETEROCHROMATIN DURING INTERPHASE AND MITOSIS, PROBABLY BY
    INTERACTING WITH HP1 (BY SIMILARITY).
-1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
    This SWISS-PROT entry is copyright. It is produced through a collaboration
    between the Swiss Institute of Bioinformatics and the EMBL outstation
    the European Bioinformatics Institute. There are no restrictions on its
    use by non-profit institutions as long as its content is in no way
    modified and this statement is not removed. Usage by and for commercial
    entities requires a license agreement (See http://www.isb-sib.ch/announce/
    or send an email to license@isb-sib.ch).
    EMBL; D64059; BAA10936.1;
    DNA repair; Nuclear protein; DNA-binding; Helicase.
    NON_TER 1
    FT DOMAIN 251 256 POLY-SER.
    FT DOMAIN 304 307 POLY-SER.
    FT DOMAIN 360 367 POLY-ASP.
    FT NON_TER 527 527
    FT SEQUENCE 527 AA; 59258 MW; ABF4B10C086D638 CRC64;
    Query Match 3.7%; Score 167.5; DB 1; Length 527;
    Best Local Similarity 20.8%; Pred. No. 0.096;
    Matches 108; Conservative 83; Mismatches 206; Indels 121; Gaps 20;
    QY 221 RMLGTRATGNTAGTGTGVARFTPSHLNKNKSCRTSLCYSEDGQELVSYSDIYLF 280
    Db 86 RQRTGR 129
    QY 281 DP-----KDTARELA-----TPSAEREELRQPVKRLRGDWSDTGPRAPESERE 330
    Db 130 SKGVKNKNDTDEAKGKKWKDKCKEEL-SDSVDRLPVKGSCDS-----EDKTK 183
    QY 331 RDGQSPNVSLMQMSDMLSRWFEEAEVAOSNRGRSRPRGTSQSDISLTPVPSP 390
    Db 184 RN-----RVSLREK-----KQSLPAKSS-----GKRPECSSDTERSVKGECCDT 225
    QY 391 DLEVSETAMEVDTPAEQFLQSTSTWSAQAHSTSPSPESHSTPLLSPPDSEQSRV 450
    Db 226 DKRVRIDLR-----ERRSSNKRSTKRVKSGSSDAEG-----SSEDAKKQKK 273
    QY 451 SGHHTHQDNNKLSKPGTGPVLSLHYSTEGTTSTIKLFTDEWSSIASSSRGIG 510
    Db 274 SA-----KKNSNTKERKK-----SLRATTTKRQADITSSSDIGDDQNSA 317
    QY 511 SHCKSEGESFVPSQSVOPPE-GDSEYKAPESSEDTKYQEGVSAENPVENHINITQS 569
    Db 318 GEESDEQKIKPVTEINLVPSHTGFCQSGDEAFKSVPAVTVDVDDDDNDPENRI-----A 373
    QY 570 DKFTAKPLDSNGERNDLNRDCSGVPEESASEKAKEPETSDQSTESATNENNT 625
    Db 374 KMLLEEIKAN-----LSSDEGSSDDEPEKEGKRIKQSEETPGDDGSGNOVSESDS 427
    QY 626 ----NPEQFQ-----TRATGTPSAHEETSTRDSALQOTDSDDDPVL 664
    Db 428 DSEESKPKRYRHLRLHKLSDGESGEEKTKPEKHEKTKGRNRRKVSSESDTDFOE 487
    QY 665 PGARYRAG-----PGDRRSARVARIQEFFRRRERKE 695
    Db 488 SGVSEVSESEDEQPRTRRSKAKAELEENQSYKQKK 525
    RESULT 25
    NP14_RAT STANDARD; PRT; 704 AA.
    ID NP14_RAT
    AC P4177:
    DT 01-NOV-1995 (Rel. 32, Created)
  
```

Query Match

```

Best Local Similarity 23.2%; pred. No. 0.23;
Matches 100; Conservative 65; Mismatches 195; Indels 71; Gaps 15;

240 MVARFIPSHLNKSC---RVTSLCYSDGQGILVYSDDIYLFDPKDDTARELKTPSAE 296
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
58 LVSSVVPFTSSSLSDTIASILSES---LVSIFFSLY-----TSSDISTSVN 106
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
297 ERRELQPPVKRLRLRGWDSTQPRAPESERERDQSPNVSLMQRMQDMLSRWFEEA 356
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
107 DVES-----SYSGPNSYALSSTNAQLSSTTETDIS--SATQTS 148
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
357 SEVAQSNRGRSRPRGGTS-----QSDISTLVPVSPDLEYSETAM-E 400
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
149 PQTSSNGGGSSSPLGKSVLETTASSDITAVTSSFTTILTVSSSPKISSGSAVTS 208
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
401 VDTPAEQFLQPTST--MSQAHSSTSSPTES-----PHSTPLL---SPDPSEQRQSV 449
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
209 VGTTSDAKVEFSSSTSDVSSLLSSSTSPASSITSETLPFSSTILTSITSSVPSAEPSAT 268
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
450 AGSHHTHQSDDNNEKLSPKPTGCEPVLSHYSTGCTTTSTIKLFTDEWSIASSSRGI 509
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
269 SSVSSEASSSSTSSVSSEAPLATSSVVS--SEAPSSSVVSSEAPSSSTSSVSSE--I 324
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
510 GSHCKGEGOBESFVQSSVQPPGEGDSETKAPBESSEDVIKYQGVSAENPVENHINIQS 569
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
325 SSITSSSVSSEAPLATSSVVSSEAPSSSTSS--SVSSEISSTSSSVSSEAPL-----ATSS 378
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
570 DKFTAPLQNSGERNDLNRSCVGPPEESASKEAPESTSDQTEASATENNNTPEP 629
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
379 VVSSEAPSSSTSSVSSEAPSSSTSSVSSEAPSSSTSS--VSSEISSTKSSVSSEVSSAT 436
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
630 QFOTEATGPSA 640
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
437 SSLVSSSSAPSA 447
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 27
ID      ATRX_CAEEL
ID      ATRX_CAEEL      STANDARD;      PRT;      1359 AA.
AC      Q9U7E0; 002061;
DT      15-JUN-2002 (Rel. 41, Created)
DT      15-JUN-2002 (Rel. 41, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Transcriptional regulator ATRX homolog (X-linked nuclear protein-1).
GN      XNP-1 OR B0041.7.
OS      Caenorhabditis elegans.
OS      Caenorhabditis elegans.
OS      Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC      Rhabditidae; Peloderinae; Caenorhabditis.
OC      NCBI_TaxID=6239;
OX      [1]
RN      SEQUENCE FROM N.A., AND CHARACTERIZATION.
RP      MEDLINE=99365296; PubMed=10433961;
RA      Villard L., Fontes M., Ewbank J.J.;
RT      "Characterization of xnp-1, a Caenorhabditis elegans gene similar to
RT      the human XNP/ATR-X gene.";
RL      Gene 236:13-19(1999).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Bristol N2;
RA      Fulcon R., Wohldmann P.;
RL      Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
CC      -!- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES
CC      GENE EXPRESSION BY AFFECTING CHROMATIN (Potential).
CC      -!- SUBCELLULAR LOCATION: Nuclear.
CC      -!- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation on its
CC      European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC      or send an email to license@isb-sib.ch).

```

[illegible]

RESULT 28  
POPCV\_HUMAN  
ID PGCV\_HUMAN  
STANDARD; PRT; 3996 AA.  
AC P13611; P20754; Q9UNW5; Q13010; Q13189; Q15123;  
DT 01-JAN-1990 (Rel. 13, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DE 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Versican core protein precursor (Large fibroblast proteoglycan)  
 DE Chondroitin sulfate proteoglycan core protein 2 (PG-M) (Glial  
 GN hyaluronate-binding protein) (CHAP).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM V0).  
 RX MEDLINE=95105186; PubMed=7528742;  
 RA Naso M.F., Zimmermann D.R., Iozzo R.V.;  
 RT "Characterization of the complete genomic structure of the human  
 RL versican gene and functional analysis of its promoter.";  
 RN J. Biol. Chem. 269:32999-33008(1994).  
 RP SEQUENCE FROM N.A. (ISOFORM V1).  
 RX TISSUE=Placenta;  
 RA Zimmermann D.R., Ruoslahti E.;  
 RT "Multiple domains of the large fibroblast proteoglycan, versican.";  
 RN EMBO J. 8:2975-2981(1989).  
 RP SEQUENCE FROM N.A. (ISOFORM V2).  
 RX TISSUE=Glial tumor;  
 RA Dours-Zimmermann M.T., Zimmermann D.R.;  
 RT "A novel glycosaminoglycan attachment domain identified in two  
 RL alternative splice variants of human versican.";  
 RN J. Biol. Chem. 269:32992-32998(1994).  
 RP SEQUENCE OF 2711-3396 FROM N.A.  
 RX TISSUE=Lung fibroblast;  
 RA Krusius T., Gehlsen K.R., Ruoslahti E.;  
 RT "A fibroblast chondroitin sulfate proteoglycan core protein contains  
 RL lectin-like and growth factor-like sequences.";  
 RN J. Biol. Chem. 262:13120-13125(1987).  
 RP SEQUENCE OF 251-347 FROM N.A.  
 RX MEDLINE=93122792; PubMed=1478664;  
 RA Iozzo R.V., Naso M.F., Cannizzaro L.A., Wasmuth J.J.,  
 RT McPherson J.D.;  
 RL "Mapping of the versican proteoglycan gene (CSPG2) to the long arm of  
 RL human chromosome 5 (5q12-5q14).";  
 RN Genomics 14:845-851(1992).  
 RP SEQUENCE FROM N.A. (ISOFORM V3).  
 RX TISSUE=Brain;  
 RA MEDLINE=95181355; PubMed=7876137;  
 RA Zako M., Shinomura T., Ujita M., Ito K., Kimata K.;  
 RT "Expression of PG-M(V3), an alternatively spliced form of PG-M  
 RL without a chondroitin sulfate attachment in region in mouse and human  
 RL tissues.";  
 RN J. Biol. Chem. 270:3914-3918(1995).  
 RP SEQUENCE OF 3333-3396 FROM N.A. (ISOFORM VINT).  
 RX TISSUE=Aortic smooth muscle;  
 RA MEDLINE=99327053; PubMed=10397680;  
 RA Lemire J.M., Braun K.R., Maurel P., Kaplan E.D., Schwartz S.M.,  
 RT Wright T.N.;  
 RL "Versican/PG-M isoforms in vascular smooth muscle cells.";  
 RN Arterioscler. Thromb. Vasc. Biol. 19:1630-1639(1999).  
 RP PARTIAL SEQUENCE.  
 RX TISSUE=Brain;  
 RA MEDLINE=89174663; PubMed=2466833;  
 RA Perides G., Lane W.S., Andrews D., Dahl D., Bignami A.;  
 RT "Isolation and partial characterization of a glial  
 RL hyaluronate-binding protein.";  
 RN J. Biol. Chem. 264:5981-5987(1989).  
 RN [9]  
 RP TISSUE SPECIFICITY OF ISOFORMS.  
 RX MEDLINE=96213482; PubMed=8627343;  
 RA Paulus W., Baur I., Dours-Zimmermann M.T., Zimmermann D.R.;  
 RT "Differential expression of versican isoforms in brain tumors.";  
 RN J. Neuropathol. Exp. Neurol. 55:528-533(1996).  
 CC - FUNCTION: May play a role in intercellular signaling and in  
 CC connecting cells with the extracellular matrix. May take part in  
 CC the regulation of cell motility, growth and differentiation. Binds  
 CC hyaluronic acid.  
 CC - SURCELLULAR LOCATION: Secreted; extracellular matrix.  
 CC - ALTERNATIVE PRODUCTS: At least 5 isoforms: V0 (shown here), V1,  
 CC V2, V3 and Vint; are produced by alternative splicing.  
 CC - TISSUE SPECIFICITY: Cerebral white matter. V0 and V1 is expressed  
 CC in normal brain, gliomas, medulloblastomas, schwannomas,  
 CC neurofibromas, and meningiomas; V2 is restricted to normal brain  
 CC and gliomas; V3 is found in all these tissues except  
 CC medulloblastomas.  
 CC - DEVELOPMENTAL STAGE: Disappears after the cartilage development.  
 CC - SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.  
 CC - SIMILARITY: CONTAINS 2 LINK DOMAINS.  
 CC - SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.  
 CC - SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
 CC - SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.  
 CC - SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 DR EMBL; U16306; AAA65018.1; -  
 DR EMBL; X15998; CAA34128.1; -  
 DR EMBL; S52488; AAB24878.1; -  
 DR EMBL; U26555; AAA67565.1; -  
 DR EMBL; D32039; BAA06801.1; -  
 DR EMBL; J02814; AAA36437.1; -  
 DR EMBL; AF084545; AAD48545.1; -  
 DR PIR; S06014; S06014.  
 DR PIR; A29348; A29348.  
 DR PIR; A30358; A30358.  
 DR HSP; P01132; IEGF.  
 DR Genew; HGNC:2464; CSPG2.  
 DR MIM; 118661; -  
 DR InterPro; IPR000152; Asx\_hydroxyl.  
 DR InterPro; IPR000561; EGF-like.  
 DR InterPro; IPR000742; EGF-2.  
 DR InterPro; IPR001881; EGF-Ca.  
 DR InterPro; IPR001438; EGF-II.  
 DR InterPro; IPR003599; Ig\_MHC.  
 DR InterPro; IPR003006; Ig\_MHC.  
 DR InterPro; IPR001304; Lectin\_C.  
 DR InterPro; IPR000538; Link.  
 DR Pfam; IPR000436; Sushi\_SCR\_CCP.  
 DR Pfam; PF00008; EGF; 2.  
 DR Pfam; PF00047; Ig; 1.  
 DR Pfam; PF00059; Lectin\_c; 1.  
 DR Pfam; PF00084; sushi; 1.  
 DR Pfam; PF00193; Xlink; 2.  
 DR PRINTS; PR00010; EGFBLD.  
 DR ProDom; PD000918; Link; 2.  
 DR SMART; SM00032; CCP; 1.  
 DR SMART; SM00034; CLECT; 1.  
 DR SMART; SM00179; EGF\_CA; 1.  
 DR SMART; SM00001; EGF\_like; 1.  
 DR SMART; SM00409; IG; 1.  
 DR SMART; SM00445; LINK; 2.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 1.  
 DR PROSITE; PS00022; EGF\_1; 2.  
 DR PROSITE; PS01186; EGF\_2; 1.

DR	PROSITE	PS01187	EGF_CA; 1	DB	740	SVEM	-----TKSFDFFTLTKLSAETVRD---MEDFTA---TPGITKYDENITT	785
DR	PROSITE	PS01241	LINK; 2	QY	415	-----STMSAQA	-----HSTSSPSTESPHSTPLSS	439
DR	PROSITE	PS00615	C-TYPE_LECTIN_1; 1	DB	786	VLLAHGTLVSEATVSKWDEEDNTTSKPLES--TEPSASSKLPPALLTTVGMNGKDKDI	843	
KW	Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi; Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain; Hyaluronic acid; Alternative splicing.	1	20	QY	440	-----PDSEQRQ	-----SVEASGHHTHH---QSDNNNEKLSKPKPGTGE	474
KW	HYALURONIC ACID; ALTERNATIVE SPLICING.	1	20	DB	844	PSFTEDGADFTLIPDSTOKLEEVTDEDIAAHGKFTIRFPQPTSGIAEKSTLRSITE	903	
FT	SIGNAL	21	3396	QY	475	PVLSLHSTSTEGTTSTIKLFTDEWSSI	-----ASSRSGIGSHCKSEGE	520
FT	CHAIN	27	137	QY	904	EKVPPITSTEGQVATNEGSGALGEVEDVLSKPVSTVPQFAHTSEVEGLAFVSYSTQEP	963	
FT	DOMAIN	167	244	QY	521	-----SFVPOS	-----SVQPPGSDSTKAPESSEDEV	547
FT	DOMAIN	265	346	QY	964	TTYVDSSTHTPLSVIPKTDGVLVPSVPSDEVLPESODILVIDOTRLERATISPETMRT	1023	
FT	DOMAIN	348	1335	QY	548	TKYQEGV	-----SAENPVENHINITQSKFTAKPLDSSNGE	587
FT	DOMAIN	1336	3089	DB	1024	TKITEGTTQEEFPWKEQTAEPKVPALSSATWTAPKAVTPLDEQEGDGSAYTVEDELLTG	1083	
FT	DOMAIN	1337	3125	QY	588	-----NLDRCGVPESASERAKPEPDSQSTSTESATNENTNPEPOFOTE	634	
FT	DOMAIN	1338	3126	QY	588	-----NLDRCGVPESASERAKPEPDSQSTSTESATNENTNPEPOFOTE	634	
FT	DOMAIN	1339	3127	QY	588	-----NLDRCGVPESASERAKPEPDSQSTSTESATNENTNPEPOFOTE	634	
FT	DOMAIN	1340	3128	QY	588	-----NLDRCGVPESASERAKPEPDSQSTSTESATNENTNPEPOFOTE	634	
FT	DOMAIN	1341	3129	QY	588	-----NLDRCGVPESASERAKPEPDSQSTSTESATNENTNPEPOFOTE	634	
FT	DOMAIN	1342	3130	QY	588	-----NLDRCGVPESASERAKPEPDSQSTSTESATNENTNPEPOFOTE	634	
FT	DOMAIN	1343	3131	QY	588	-----NLDRCGVPESASERAKPEPDSQSTSTESATNENTNPEPOFOTE	634	
FT	DOMAIN	1344	3132	QY	588	-----NLDRCGVPESASERAKPEPDSQSTSTESATNENTNPEPOFOTE	634	
FT	DOMAIN	1345	3133	QY	588	-----NLDRCGVPESASERAKPEPDSQSTSTESATNENTNPEPOFOTE	634	
FT	DOMAIN	1346	3134	QY	588	-----NLDRCGVPESASERAKPEPDSQSTSTESATNENTNPEPOFOTE	634	
FT	DOMAIN	1347	3135	QY	588	-----NLDRCGVPESASERAKPEPDSQSTSTESATNENTNPEPOFOTE	634	
FT	DOMAIN	1348	3136	QY	588	-----NLDRCGVPESASERAKPEPDSQSTSTESATNENTNPEPOFOTE	634	
FT	DOMAIN	1349	3137	QY	588	-----NLDRCGVPESASERAKPEPDSQSTSTESATNENTNPEPOFOTE	634	
FT	DOMAIN	1350	3138	QY	588	-----NLDRCGVPESASERAKPEPDSQSTSTESATNENTNPEPOFOTE	634	
FT	DOMAIN	1351	3139	QY	588	-----NLDRCGVPESASERAKPEPDSQSTSTESATNENTNPEPOFOTE	634	
FT	DOMAIN	1352	3140	QY	588	-----NLDRCGVPESASERAKPEPDSQSTSTESATNENTNPEPOFOTE	634	
FT	DOMAIN	1353	3141	QY	588	-----NLDRCGVPESASERAKPEPDSQSTSTESATNENTNPEPOFOTE	634	
FT	DOMAIN	1354	3142	QY	588	-----NLDRCGVPESASERAKPEPDSQSTSTESATNENTNPEPOFOTE	634	
FT	DOMAIN	1355	3143	QY	588	-----NLDRCGVPESASERAKPEPDSQSTSTESATNENTNPEPOFOTE	634	
FT	DOMAIN	1356	3144	QY	588	-----NLDRCGVPESASERAKPEPDSQSTSTESATNENTNPEPOFOTE	634	
FT	DOMAIN	1357	3145	QY	588	-----NLDRCGVPESASERAKPEPDSQSTSTESATNENTNPEPOFOTE	634	
FT	DOMAIN	1358	3146	QY	588	-----NLDRCGVPESASERAKPEPDSQSTSTESATNENTNPEPOFOTE	634	
FT	DOMAIN	1359	3147	QY	588	-----NLDRCGVPESASERAKPEPDSQSTSTESATNENTNPEPOFOTE	634	
FT	DOMAIN	1360	3148	QY	588	-----NLDRCGVPESASERAKPEPDSQSTSTESATNENTNPEPOFOTE	634	
FT	DOMAIN	1361	3149	QY	588	-----NLDRCGVPESASERAKPEPDSQSTSTESATNENTNPEPOFOTE	634	
FT	DOMAIN	1362	3150	QY	588	-----NLDRCGVPESASERAKPEPDSQSTSTESATNENTNPEPOFOTE	634	
FT	DOMAIN	1363	3151	QY	588	-----NLDRCGVPESASERAKPEPDSQSTSTESATNENTNPEPOFOTE	634	
FT	DOMAIN	1364	3152	QY	588	-----NLDRCGVPESASERAKPEPDSQSTSTESATNENTNPEPOFOTE	634	
FT	DOMAIN	1365	3153	QY	588	-----NLDRCGVPESASERAKPEPDSQSTSTESATNENTNPEPOFOTE	634	
FT	DOMAIN	1366	3154	QY	588	-----NLDRCGVPESASERAKPEPDSQSTSTESATNENTNPEPOFOTE	634	
FT	DOMAIN	1367	3155	QY	588	-----NLDRCGVPESASERAKPEPDSQSTSTESATNENTNPEPOFOTE	634	
FT	DOMAIN	1368	3156	QY	588	-----NLDRCGVPESASERAKPEPDSQSTSTESATNENTNPEPOFOTE	634	
FT	DOMAIN	1369	3157	QY	588	-----NLDRCGVPESASERAKPEPDSQSTSTESATNENTNPEPOFOTE	634	
FT	DOMAIN	1370	3158	QY	588	-----NLDRCGVPESASERAKPEPDSQSTSTESATNENTNPEPOFOTE		

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; L13773; AAA58360.1; -;  
 DR EMBL; L25050; AAA36642.1; -;  
 DR Genew; HGNC:7135; MLLT2.  
 DR MIM; 159557; -;  
 KW Nuclear protein; Chromosomal translocation; Proto-oncogene.  
 FT DOMAIN 483 492 POLY-SER.  
 FT DOMAIN 835 843 POLY-SER.  
 FT DOMAIN 866 869 POLY-PRO.  
 FT DOMAIN 871 874 POLY-SER.  
 FT CONFLICT 46 46 K -> R (IN REF. 2).  
 FT CONFLICT 624 624 E -> G (IN REF. 2).  
 FT CONFLICT 899 905 SASSTKS -> VFAPRV (IN REF. 2).  
 FT CONFLICT 928 929 EH -> AD (IN REF. 2).  
 FT CONFLICT 999 999 I -> N (IN REF. 2).  
 FT CONFLICT 1096 1096 A -> AR (IN REF. 2).  
 FT CONFLICT 1140 1140 N -> I (IN REF. 2).  
 FT CONFLICT 1177 1210 STNVCTIALNSLVDLVHYTRQGFQQLQELTKTP -> RQ  
 MCAPFESTAVWMTWCTIHDRVFSYKN (IN REF. 2).  
 FT SQ SEQUENCE 1210 AA; 131421 MW; F0E334DF8CF2FF04 CRC64;

Query Match 3.7%; Score 165.5; DB 1; Length 1210;  
 Best Local Similarity 20.4%; Pred. No. 0.32;  
 Matches 125; Conservative 80; Mismatches 218; Indels 189; Gaps 26;

QY 279 LFDPKDDTARELTPTSAEERREELRPPVKRLRLGDSVTPG-----RARPESERED 332  
 Db 263 LVAPAPPSPQTFPPPLPSKSAVMOQKPTAYVRPM-DGQDQAPSESPKLPEDYRQQT 321  
 QY 333 GEQS-----PNVSLMORMS-----DMLSRW----- 352  
 Db 322 FEKTLKVPKAKLTKLKPQSVQVEGTYNVHCVEILKEMTHSPPLTAHTPTSTAE 381  
 QY 353 -----FEASVAQSNRGRSRPRGTSQSDISTLTPTVSPDLEVSSETAMEVD 402  
 Db 382 PSKFFPTKDSQHVSVTONQYDTSKTHSNQOQTSSM-----LEDDLQSDSE---D 434  
 QY 403 TPAEQFLO--PSTSSSTMSAQAHSTSPSTPSTPLSSPDSPQROSVASGHHHTHQSDN 461  
 Db 435 SDSEQTEPKPPSSAPPSPQSLPEVASAHS-----SSAESTSDSDSS-----SDS 483  
 QY 462 NNEKLSPPGCTGEPVLSHYSTEGTTTITKINFTEW-----SSIASSSRGIGS----- 511  
 Db 484 ESESSSDSENEPLETAPPEPTTNKWL-----DNMLTKVSOAPAPPEGRSTPEPRR 540  
 QY 512 HCKSGQESFVPO--SSVQPPGDSSTKAPESSE-----DVTKYQGVSAENPVENHIN 565  
 Db 541 HPESKSSDSATSQHSSESKDPPPKSSSKAPRAPPEAPHGKRKSCQSPAQEPPOQ-- 598  
 QY 566 ITQSDKFTAKPL--DSNSGERNDLMDRCGV-----PEESASS 602  
 Db 599 -TVGTOKPKVKASARAGSRTSLQGEREPCLLYGSDQTSKDKPKVKTKGRPRAAASN 657  
 QY 603 E-KAKEPETSQTES---ATNENNTNPEPOQ-----TEATGFSAEHE 643  
 Db 658 EPKPAVPPSEKKHKHSLPAPSKALSGPEAKDNVEDRTPHEFALVPLTQSGP-PHSG 716  
 QY 644 TSTRDSALQDT-----DSDDDDPVLIP----- 665  
 Db 717 SGRSTSGCRQAVVQEDSRKDRPLPLRDTKLLSPDRTPPPQSLMVKYITLDLSRIQP 776  
 QY 666 ---GARYRAG-----PGDRRAVARIQEFFFF--RKERKEMEELDTNIRPLVKMYKG 715  
 Db 777 PGKSGRQKAEADKOPAGKAGKHSEKSSSDSSKLAKRKKEARDCDNKKIRLEKEIKSQ 836  
 QY 716 HRNSRTMIKEAN 727  
 Db 837 SSSSSSSSHKESS 848

## RESULT 30

EBAL\_PLAFC STANDARD; PRT; 1435 AA.  
 ID EBAL\_PLAFC  
 AC P19214;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE Erythrocyte-binding antigen EBA-175.  
 OS Plasmodium falciparum (isolate Camp / Malaysia).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.  
 OX NCBI\_TaxID=5835;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 EX MEDLINE=90377299; PubMed=2204835;  
 RA Sim B.K.L.;  
 RT "Sequence conservation of a functional domain of erythrocyte binding  
 antigen 175 in plasmodium falciparum."  
 RL Mol. Biochem. Parasitol. 41:293-296(1990).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; X52524; CAA36756.1; -;  
 DR EMBL; X52524; CAA36756.1; -;  
 DR PIR; S11561; S11561.  
 KW Antigen.  
 FT DOMAIN 159 1104 ESSENTIAL FOR BINDING TO  
 FT VARIANT 1031 1031 ERYTHROCYTES.  
 FT SEQUENCE 1435 AA; 167389 MW; 32A4309021B1C3D6 CRC64;

Query Match 3.7%; Score 165.5; DB 1; Length 1435;  
 Best Local Similarity 18.1%; Pred. No. 0.39;  
 Matches 188; Conservative 148; Mismatches 330; Indels 375; Gaps 53;

QY 36 EFTORLKLKATLVHVDGCVNTIC-----WNDTGEYILSGDDTKLVISN 79  
 Db 290 EFERKL-WEAMUSEHKNNIN--CKNIPOBELQITOWIKWH--GEFLERDNRSL 341  
 QY 80 PYSRKVLTTIRSG-----HRANIFSAK-----LPCNDK---QI 111  
 Db 342 PKSKCKNNLTLYEACEKCDPCMKYRDWIRSKFEWHTLSKEYETOKVPKNAENLYKI 401  
 QY 112 VSCSGDGVIFYTNVQDAETNRQCOFTCHYTGYEIMTPNDPPTFLSCGEGDTVRW--- 168  
 Db 402 SENKNDKAVSLLNNDCAEYSKYCD--CKHTTTL-VKSVLN-----GNDNTKEKRE 450  
 QY 169 ---FTRIKTSCTKE-----DCKDILLNCRATSAICPPPIPYLAVGSDSSV 216  
 Db 451 HDLDDDFSKGCKNSVDNTNFKWECKNPYIL-----STKDCVPPRRQELCLGNID--- 502  
 QY 217 RIYDRML-----  
 Db 503 RIYDKNLMLKEHILAIAYESRLKRYKKNODKEVCKIINKTFADINDIIGDYNWD 562  
 QY 232 YAGRTTGMV---ARFIPSHLNKSCR-----TSLCYSDG 265  
 Db 563 LSNKRLVGKINTNSKIY--HRNKNDKLFDRDWMVKIKDVMNVISWVFKDKTVCKEDDI 620  
 QY 266 QEILVSYSDYIYLPKDDTARELTPTSAEERREELRPPVK-----RLRLGDSV 318  
 Db 621 ENI-----POFRRFSEWGGDDYCDQCKYKMTLKVCKEKEKPEDDCKNCKSNYKWLK 675  
 QY 319 TGRPARPESERERDGEOSPVSMLQRMDSM-----LSRWFEASEVAQSNRGRSRPRG 373  
 Db 676 KKEYNKQAKQYQYQYQYQYQYQYSEFKSIPKPEVLYKSEKSNLNFEDFEKEE----- 730  
 QY 374 GTSQSDISTLTPTVSPDLEVSSETAMEVDTPAQFLQPSISSTMSAQAHSTSP--TESP 431

Db 731 --LHSDYKNCMTM--CP--EVKDPISIRNNEQTQSAVPEENTEIAHRTETESISEGP 784  
QY 432 HSTPLLSFDSQROSVASGHHHQDNNNEKLSKPKGPGTGEVLSLHYSTEGTTSTI 491  
Db 785 -----KGNEQKERDDSLSKTSVSPENSRP-----ETDAKDTISNL 819  
QY 492 -----KLANTEWSSIASRRGIGSHCKSEGOEFSFVPQSSVQ 529  
Db 820 LKLKGDVLSMPKAVIGSPNDINVTQEGDNIS-----GVNSKPLSD-----DVR 865  
QY 530 PPGEDSETKAPESSEDVTKYQGVSAENPVENHINIQSDKFTAKPLDSNSG-----ER 584  
Db 866 PDKKELEDQNSDESETVYVNH-----ISKSPSINNDDSGSGSVATVSESSSNTGLSIDDDR 922  
QY 585 NDLNDRSCGVPESASSKAKPEPSTQSTSTES-----ATNENNTNPEPQFQTEATG 637  
Db 923 NGDTFFVKT-----QDTANTEDVIRKENADKDEKGADEERHSTSESLSPSEKMLTDNEG 978  
QY 638 PSA--HEE-----TSTRDSA-----LQDT-----DD-----656  
Db 979 GNSLNHEEVKEHTSNSDNVQSGGIYVMNVKELKDTLENPSSLDGKAHSELSEPNLS 1038  
QY 657 SDDPVLIPGARYAGGDRSARVATQEFRR--RKERKEMEELDTLIRPLVKNVYK 714  
Db 1039 SQQDMSNTP-----GPLDNTS-----EETTERISNNEKYNREDERTLTKKEYEDIVLK 1087  
QY 715 GHRNSR-----TMKEANFWGANFVNGSGDCGHFIWDRHTAHLMLLEAD 760  
Db 1088 SHMNRSDGELYDENSIDLTVNDESE--DAEAKMGNDTSEM--SHNSQHR--IESD 1139  
QY 761 -----NHVNCLOPHPEPILASSGIDYDI-----KIWSPLEERSIFRN-- 798  
Db 1140 QQKNDMTVGLGTHYVQNEIS-----VPVTGEIDKLESKESKI-HKAEBERLSHTD 1192  
QY 799 --RKLADDEVITRNELMLETRN 818  
Db 1193 IHKINPEDNRNNTLHLKDIRN 1213  
RESULT 31  
ANK2\_HUMAN STANDARD; PRT; 3924 AA.  
AC Q01484; Q01485;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Ankyrin 2 (Brain ankyrin) (Ankyrin B) (Ankyrin, nonerythroid).  
GN ANK2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP TISSUE=Brain stem;  
RC MEDLINE=91302466; PubMed=1830053;  
RA Otto E., Kunimoto M., McLaughlin T., Bennett V.;  
RT "Isolation and characterization of cDNAs encoding human brain  
RT ankyrins reveal a family of alternatively spliced genes.";  
RL J. Cell Biol. 114:241-253(1991).  
RN [2]  
RP REVISIONS.  
RP Carpenter S.;  
RA Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RC TISSUE=Brain stem;  
RX MEDLINE=94075409; PubMed=8253844;  
RA Chan W., Kordeli E., Bennett V.;  
RT "440-kD ankyrinB: structure of the major developmentally regulated  
RT domain and ankyrinB: structure of the major developmentally regulated  
RL J. Cell Biol. 123:1463-1473(1993).

[4]  
RN SEQUENCE OF 463-495 FROM N.A.  
RP MEDLINE=92009921; PubMed=1833308;  
RA Tse W.T., Menninger J.C., Yang-Peng T.L., Francke U., Sahr K.E.,  
RA Lux S.E., Ward D.C., Forget B.G.;  
RT "Isolation and chromosomal localization of a novel nonerythroid  
RT ankyrin gene.";  
RL Genomics 10:858-866(1991).  
CC -1- FUNCTION: Attach integral membrane proteins to cytoskeletal  
CC elements. Also bind to cytoskeletal proteins.  
CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1 (SHOWN HERE), 2 AND 3; ARE  
CC PRODUCED BY ALTERNATIVE SPLICING.  
CC -1- TISSUE SPECIFICITY: PLASMA MEMBRANE OF NEURONS AS WELL AS GLIAL  
CC CELLS THROUGHOUT THE BRAIN.  
CC -1- PTM: PHOSPHORYLATED AT MULTIPLE SITES BY DIFFERENT PROTEIN KINASES  
CC AND EACH PHOSPHORYLATION EVENT REGULATES THE PROTEIN'S STRUCTURE  
CC AND FUNCTION (POTENTIAL).  
CC -1- SIMILARITY: CONTAINS 23 ANK REPEATS.  
CC -1- SIMILARITY: CONTAINS 1 DEATH DOMAIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X56957; CAA40278.1; -;  
CC EMBL; X56958; CAA40279.2; -;  
CC EMBL; Z26634; CAB42644.1; -;  
CC EMBL; M37123; AAA62828.1; -;  
CC PIR; S14533; S14533.  
CC PIR; A39643; A39643.  
CC PIR; B39643; B39643.  
CC PIR; S14569; S14569.  
CC HSP; P42771; IDC2.  
CC Genew; HGNC:493; ANK2.  
CC MIM; 106410; -;  
CC InterPro; IPR002110; ANK.  
CC InterPro; IPR000488; Death.  
CC InterPro; IPR000906; ZUS.  
CC Pfam; PF0023; ank; 24.  
CC Pfam; PF00531; death; 1.  
CC Pfam; PF00791; ZUS; 1.  
CC PRINTS; PR01415; ANKYRIN.  
CC SMART; SM00248; ANK; 21.  
CC SMART; SM00005; DEATH; 1.  
CC SMART; SM00218; ZUS; 1.  
CC PROSITE; PS50088; ANK\_REPEAT; 20.  
CC PROSITE; PS50297; ANK\_REPEAT\_REGION; 1.  
CC PROSITE; PS50017; DEATH\_DOMAIN; 1.  
KW Cytoskeleton; Alternative splicing; Repeat; ANK repeat;  
KW Phosphorylation.  
FT REPEAT 63 92 ANK 1.  
FT REPEAT 96 125 ANK 2.  
FT REPEAT 129 158 ANK 3.  
FT REPEAT 162 191 ANK 4.  
FT REPEAT 193 220 ANK 5.  
FT REPEAT 232 261 ANK 6.  
FT REPEAT 265 294 ANK 7.  
FT REPEAT 298 327 ANK 8.  
FT REPEAT 331 360 ANK 9.  
FT REPEAT 364 393 ANK 10.  
FT REPEAT 397 426 ANK 11.  
FT REPEAT 430 459 ANK 12.  
FT REPEAT 463 492 ANK 13.  
FT REPEAT 496 525 ANK 14.  
FT REPEAT 529 558 ANK 15.  
FT REPEAT 562 591 ANK 16.  
FT REPEAT 595 624 ANK 17.  
FT REPEAT 628 657 ANK 18.  
FT REPEAT 661 690 ANK 19.



FT REPEAT 694 723 ANK 20.  
 FT REPEAT 727 756 ANK 21.  
 FT REPEAT 760 789 ANK 22.  
 FT REPEAT 793 822 ANK 23.  
 FT DOMAIN 1773 1950 REPEAT-RICH REGION.  
 FT REPEAT 1773 1784 REPEAT A.  
 FT REPEAT 1785 1796 REPEAT A.  
 FT REPEAT 1797 1808 REPEAT A.  
 FT REPEAT 1809 1820 REPEAT A.  
 FT REPEAT 1821 1832 REPEAT A.  
 FT REPEAT 1833 1844 REPEAT A.  
 FT REPEAT 1845 1856 REPEAT A.  
 FT REPEAT 1857 1867 REPEAT A.  
 FT REPEAT 1868 1879 REPEAT A.  
 FT REPEAT 1880 1891 REPEAT A.  
 FT REPEAT 1892 1902 REPEAT A.  
 FT REPEAT 1903 1914 REPEAT A.  
 FT REPEAT 1915 1926 REPEAT A.  
 FT REPEAT 1927 1938 REPEAT A.  
 FT REPEAT 1939 1950 REPEAT A.  
 FT DOMAIN 3536 3620 DEATH.  
 FT VARSPLIC 1039 1039 Q -> QELGKLHLPTAPPPLNEGSLVSRILQLGPPGCK  
 (IN ISOFORM 2).  
 FT VARSPLIC 1444 3528 MISSING (IN ISOFORM 2 AND ISOFORM 3).  
 FT CONFLICT 475 476 GQ -> PE (IN REF. 4).  
 FT CONFLICT 971 971 I -> S (IN REF. 1).  
 FT CONFLICT 3581 3582 QY -> HA (IN REF. 1).  
 FT CONFLICT 3586 3586 I -> Y (IN REF. 1).  
 SQ SEQUENCE 3924 AA; 430337 MW; 52AC496C428E29D2 CRC64;

Query Match 3.7%; Score 165.5; DB 1; Length 3924;  
 Best Local Similarity 18.4%; Pred. No. 1.2;  
 Matches 111; Conservative 100; Mismatches 265; Indels 127; Gaps 22;

QY 292 TPSAEERRELRQPPVKRLRLGDSWDTGPRAPESERERDQGO-----SPNVSLMQRM 345  
 DB 2507 TPSESEVSEVTKT-----DVSTPKPAVTHECAEEDDSNGKKRRTPEENPKMV 2559  
 QY 346 SDMLSRWFEASEVAQSNRGR-----SRPGTQSQDSTLIP 384  
 DB 2560 TKI--KMFDELEQAKKDYKEPQESSSSSDPDACSDVDVDPKHTGSGDESGVP 2617  
 QY 385 TVPSSPDLEVSETAMEVDYPAQFLOPSTSSVMSAQAHTSSPTSPHSTLLSSPDSEQ 444  
 DB 2618 VLVTSESRKVVSSS-ESEPELAQLKKGADSGLLPEPVIRVQPPSPPLPSMDSNSSFEEVQ 2676  
 QY 445 RQSVASGHHTHQSDNNKLSKPKPGTGPVLSLH-----YSTGTTSTIKLP 495  
 DB 2677 FQPV-VSKQYTFKMNEDTQE-----EPGRSEKDSHSLAEDRHAVSTAEADRSYDKLR 2731  
 QY 496 TDWSSIASRRGIGSHCKSEGESEFVQSSVQPPGDSSTKAPESSEDEVTKQEGVS 555  
 DB 2732 DTQPKICD-----GHGC-----EAMSPSSARPVSSGLQSPGTGDDVDEQPVLYKESLA 2780  
 QY 556 AENPVENHINIQSDKFTAKPLDSSNGRNDLNLDRSCGVPESEASSEKAKPETSDQTS 615  
 DB 2781 LQG-----THEKDEGEELDVSAESQADCPSEFSFSSSLPHCLVSEKELDEDISATSS 2837  
 QY 616 TE-----SATNNTNPEQFOTATGPSAHEETSTRDSALQTDSDO--DPVLIPGARY 669  
 DB 2838 IOKTEVTKTDEFENLPLKDCPSQSSITQTDRFSMDVPVSDLAENDEIYDQIT--SPY 2895  
 QY 670 RAGPDRRSAVARIQEFFRRKERKEME-----ELDTNLRPLVKMYKGRHNS 719  
 DB 2896 ENVPS-----QSFSSSESTQTDANHTTSPHSSEVTSVITTSVEDVWVASSSSG 2946  
 QY 720 RTMTKEANFWGANFVMSGDCGHIFIDWRHTAEHLMLLEADNHVNCLOPHFPDILAS- 778  
 DB 2947 TVLSKESNFEQDQIKM-----ESQLESTLWQKSDS-----VSSSEFETMSAT 2989  
 QY 779 -----SGIDYDKINSPL-EESRIFNRKLADEVITRNLMLLE-ETNTI--T 821  
 DB 2990 TTVVGEIQSVIITKTDVSDSSEIRDEDEAFARVKEEQKIFGLMVDROQSOGTTPDT 3049

QY 822 VPA 824  
 DB 3050 TPA 3052

## RESULT 32

ALAL\_CANAL  
 ID ALAL\_CANAL STANDARD; PRT; 1419 AA.  
 AC O13368;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Agglutinin-like protein ALAI precursor (Agglutinin-like adhesin).  
 GN ALAI OR ALS5.  
 OS Candida albicans (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
 OX NCBI\_TaxID=5476;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98053977; PubMed=9393828;  
 RA Gaur N.K., Klotz S.A.;  
 RT "Expression, cloning, and characterization of a Candida albicans  
 gene, ALAI, that confers adherence properties upon Saccharomycetes  
 cerevisiae for extracellular matrix proteins.";  
 RL Infect. Immun. 65:5289-5294(1997).  
 CC -I- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.  
 CC -I- PFM: N-GLYCOSYLATED AND O-GLYCOSYLATED (POTENTIAL).  
 CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AF025429; AAB88883.1;  
 CC Cell adhesion; Glycoprotein; Repeat; Signal.  
 KW SIGNAL 17  
 FT CHAIN 18 1419 AGGLUTININ-LIKE PROTEIN ALAI.  
 FT DOMAIN 399 404 POLY-THR.  
 FT DOMAIN 408 418 POLY-THR.  
 FT DOMAIN 437 441 POLY-THR.  
 FT DOMAIN 673 676 POLY-SER.  
 FT DOMAIN 687 690 POLY-SER.  
 FT DOMAIN 700 703 POLY-SER.  
 FT DOMAIN 719 724 POLY-SER.  
 FT DOMAIN 749 752 POLY-SER.  
 FT DOMAIN 787 791 POLY-SER.  
 FT DOMAIN 869 872 POLY-SER.  
 FT DOMAIN 875 883 POLY-SER.  
 FT DOMAIN 901 911 POLY-SER.  
 FT DOMAIN 1216 1221 POLY-SER.  
 FT CARBOHYD 665 665 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 919 919 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1301 1301 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1326 1326 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 1419 AA; 149635 MW; 249F33F688A9D5B6 CRC64;

## Query Match

Best Local Similarity 3.6%; Score 164.5; DB 1; Length 1419;  
 Matches 129; Conservative 109; Mismatches 237; Indels 211; Gaps 29;

QY 50 HDGCYNTI-CWND---TGEYILSGSDDT-KLVISNPYSRKVLTT----- 88  
 DB 542 HNPVTWTKFWSSEVATETITNKPEGTDSDVTKPEYPTVTTFEWSSEVATETITNG 601  
 QY 89 -----IRSGHRANIFSAKFLPCTNDKQIVSCGDGVIFVTNVEQDAENRQCQFTCH 140  
 DB 602 PEGTDSVIVREPHNPVTTFTEF-----WSES 627



RN RP RX RA RT RT RL CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC

CC CC CC CC CC CC CC DR DR DR DR TW TW TW

TTTTTT

FF F F F F F F F F F F

TTTQ

Math

--  
MBL; Z34289; CAA84063.1; --  
MBL; D21262; BAA04803.1; --  
enew; HGNC:15608; NOLCL.  
IN: 602394; --  
nuclear protein; Repeat; GTP-binding; ATP-binding;  
alternative splicing.

Match	3.6%	Score 162.5;	DB 1;	Length 699;
Local Similarity	20.1%;	Pred. No. 0.24;		
Gaps	98;	Conservative 86;	Mismatches 210;	Indels 93; Gaps 20;
47 SHLNKSCRVTSLCYSEDGEIIVSYSDVIYLFDPKDITARELKTPTSAEREREELRPP	306   :   :	:   :	:   :	:   :
28 SEVANKFAKATCAQODANASSLD-----IIYSF-----WLAKSVKPERKLQANGPV	74   :   :	:   :	:   :	:   :
07 VKRLRLRGDSWDTGPRARPESEREDGEQPNSYSL-MQRMSDWLSRWEEASEVAQNRSRG	365   : :   :	:       :	: :   :	:       :

Match            3.68; Score 162.5; DB 1; Length 699;  
Local Similarity 20.18; Pred. No. 0.24;  
Res 98; Conservative 86; Mismatches 210; Indels 93; Gaps 20;

47 SHLNKSCRVTSLCYSEDGQELVSYSDVIYLFDPKDITARELKTPTSAERERELRPP 306  
| : | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
28 SEVANFKFAKATCAQTQDANASSLLD-----IYSF-----WLAKSAPVKRLQANGPV 74  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

07 VKRLRLRGDSWDTGPRARPESERERDGEQSPNYSL-MQRMSDMLSRWEESAQAQSNRG 365  
| : : : ||| : | | | | : : : | : : : | : | : | : | : | : | : | : | : |

```

CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z49198; CAA89059.1; -
CC EMBL; U20618; AAB64526.1; -
CC SGD; S0004322; CHS5;
CC InterPro; IPR001357; BRCT.
CC InterPro; IPR003961; FN_III.
CC Pfam; PF00041; fn3; 1.
CC Pfam; PF00533; BRCT; 1.
CC SMART; SM00292; BRCT; 1.
CC SMART; SM00060; FN3; 1.
CC PROSITE; PS01172; BRCT; UNKNOWN_1.
CC Repeat.
CC KW DOMAIN 76 166 FIBRONECTIN TYPE-III.
CC FT SEQUENCE 671 AA; 73638 MW; FA92741B862814C2 CRC64;
CC SQ

```

Query Match 3.6%; Score 161.5; DB 1; Length 671;  
Best Local Similarity 20.9%; Pred. No. 0.25;  
Matches 84; Conservative 59; Mismatches 141; Indels 117; Gaps 14;

```

QY 283 KDTARELKTSAEERREELRQPPVKRLRLRGDSMTGPRARPESEERERDGEQSPNVSLM 342
DB 351 ENETTTETVNPVS---RSLKSEPVGTPIENKADSAEAVVEEPNEVAESSNEAT 406
QY 343 QRMS-----DMLSRWFEEASEVAQSN---RGRGRPRG-----GTSQSDISTLP 384
DB 407 GKSEDTTHNEQADNGFVQTEVEAENNIITESAGENNEPADDAAAMEFGRPEAEIET-- 464
QY 385 TVPSSFDLEVSETAMEVDTPAEQFQLPSTSTMSQAQHSSTPSTPSTPLSPDSEQ 444
DB 465 -----PEVNESIEDANEPAEDSNPEVDSNPKVK---DSNKPVEDS-----NKPVEDS 509
QY 445 ROSVEASGHTTHQSDNNNEKLSRPGGTGEPVLSLHYSTEGTTTSTIKLNFIDEWSSIAS 504
DB 510 NKPVEDSNKPVEDANEPEVDTSEPVEDAGEPVQE-----TNEFTTIDIA 552
QY 505 SSRGIGSHCKSEGEESFVQSSVQPPGDSSETKAPESSESDVTKYQGVSAENPVENHI 564
DB 553 SPR-----HQEDIELEAE---PKDATESVAVEPSNEDVKPEEKGEAEADDINN-- 598
QY 565 NITQSKDKTAKPLDSNGERNDLNDRCGVPESASSEKAKEPETSDOTSTESATNENN 624
DB 599 -----VSKEASGE-----STHQKTEASASLES 623
QY 625 TNPEQFQTEATGPSAHEETSTRD-----SALQDTDDSDDD 660
DB 624 AVTEEQETEA-----EVTDDVLSTKEAKKNTGNSN 657

```

RESULT 36  
HET1\_PODAN STANDARD; PRT; 1356 AA.  
ID HET1\_PODAN  
AC Q00808;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Vegetative incompatibility protein HET-E-1.  
GN HET-E1.  
OS Podospora anserina.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Sordariales; Lasiosphaeriaceae; Podospora.  
OX NCBI\_TaxID=5145;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96009891; PubMed=7557402;  
RA Saupe S., Turcq B., Begueret J.;

RT "A gene responsible for vegetative incompatibility in the fungus  
 RT Podospora anserina encodes a protein with a GTP-binding motif and G  
 RL beta homologous domain."  
 CC beta 162:135-139(1995).  
 CC -!- FUNCTION: RESPONSIBLE FOR VEGETATIVE INCOMPATIBILITY THROUGH  
 CC SPECIFIC INTERACTIONS WITH DIFFERENT ALLELES OF THE UNLINKED GENE,  
 CC HET-C.  
 CC -!- SIMILARITY: CONTAINS 10 WD REPEATS (TRP-ASP DOMAINS).  
 CC -!- SIMILARITY: CONTAINS 1 NACHT DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; L28125; AAA85775.1; -.  
 CC InterPro; IPR001680; WD40.  
 CC Pfam; PF00400; WD40; 10.  
 CC PRINTS; PR00320; GPROTEINRPT.  
 CC ProDom; PD000018; WD40; 10.  
 CC SMART; SM00320; WD40; 10.  
 CC PROSITE; PS00837; NACHT; 1.  
 CC PROSITE; PS00678; WD\_REPEATS\_1; 10.  
 CC PROSITE; PS50082; WD\_REPEATS\_2; 10.  
 CC PROSITE; PS50294; WD\_REPEATS\_REGION; 1.  
 CC GTP-binding; Repeat; WD repeat.  
 CC DOMAIN 294 629  
 CC NP\_BIND 300 307 GTP (POTENTIAL).  
 CC REPEAT 839 869 WD 1.  
 CC REPEAT 881 911 WD 2.  
 CC REPEAT 923 953 WD 3.  
 CC REPEAT 965 995 WD 4.  
 CC REPEAT 1007 1037 WD 5.  
 CC REPEAT 1049 1079 WD 6.  
 CC REPEAT 1091 1121 WD 7.  
 CC REPEAT 1133 1163 WD 8.  
 CC REPEAT 1175 1205 WD 9.  
 CC REPEAT 1217 1247 WD 10.  
 CC SEQUENCE 1356 AA; 149765 MW; 965FB319844E0651 CRC64;

Query Match 3.5%; Score 160.5; DB 1; Length 1356;  
 Best Local Similarity 20.9%; Pred. No. 0.64;  
 Matches 97; Conservative 65; Mismatches 182; Indels 121; Gaps 21;  
 QY 46 TLNVHDCVNTICWNTGEVILSGDDTKLVISNPYSRKVLTTIRSGHRANIFSAKFLPC 105  
 Db 920 TLEHGGRVQSVAFSPDQGVASGDDHTTKIWDASGCTQTL-EGHGSVLVAFSP- 977  
 QY 106 TNDKQIVSCGSGGVIFVTNVQDAETNRQCFTC-----HYGTYEIMTVPNPDTFLS 159  
 Db 978 -DQQRVASGSDTKIWDATSG-----TCTQTEHGGSVWVAFSP-DQQRVAS 1026  
 QY 160 CGEDGVWFDFRIKTSCKEDKDDILNCRRAATVAICPPIPYLVAGCSDSSVRIY 219  
 Db 1027 GSDDKTIKIWDATSGT-CTQ-----TLEHGGRVQSVAFSPDQGVASGDDHTTKI 1078  
 QY 220 DRRLMGLTATNGVAGTGTGMVARETPSHLNKSCRWTSCLYSEDGEIILVSSDYIYL 279  
 Db 1079 D-AVSGT-CTQTEHGDS-----VWVAFSPDQGVASGSDGTIKI 1119  
 QY 280 FPKDDTARELTPSAERREELRQPPVKRLRLGSDTGPRAPESEERDEQSPNV 339  
 Db 1120 WDAASGTC-----TQTEHGGRVWHS-VAFSPDQGVASGSDGTI 1159  
 QY 340 SLWQRMSDMLSRWFEASEVAQSNRGRSRPRGGTSQSDISTLPTVPSPDLEVSETAM 399  
 Db 1160 KI-----WDAASGCTQTEG-----HGGVWQS-----VAFSPD----- 1188  
 QY 400 EVDTPAEQFLQSTSTMSA--QAHTSSTPSTPHSTPLLS---SPDSQRQSVASGHH 454  
 Db 400 EVDTPAEQFLQSTSTMSA--QAHTSSTPSTPHSTPLLS---SPDSQRQSVASGHH 454

Db 1189 -----GQVASSGSDTKIWDATSGTCTQTEHGGRVQSVAFSPDQGR-----VASG-- 1237  
 QY 455 THHSDNNNEKLSPKPGTGEVLSLHYSSTGTTSTIKLNFTEW 499  
 Db 1238 ---SSDNTIKIWDATSGTCTQTLNV-----GSTATCLSFDTYNAY 1274  
 RESULT 37  
 ANK3\_HUMAN  
 ID ANK3\_HUMAN STANDARD; PRT; 4377 AA.  
 AC Q12955;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Ankyrin 3 (ANK-3) (Ankyrin G).  
 GN ANK3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain stem;  
 RX MEDLINE=95138209; PubMed=7836469;  
 RA Kordeli E., Lambert S., Bennett V.;  
 RT "Ankyrin. A new ankyrin gene with neural-specific isoforms localized  
 RT at the axonal initial segment and node of Ranvier.";  
 RL J. Biol. Chem. 270:2352-2359(1995).  
 CC -!- FUNCTION: Membrane-cytoskeleton linker.  
 CC -!- ALTERNATIVE PRODUCTS: A number of isoforms are produced by  
 CC alternative splicing.  
 CC -!- TISSUE SPECIFICITY: Expressed in brain and other tissues.  
 CC -!- SIMILARITY: CONTAINS 23 ANK REPEATS.  
 CC -!- SIMILARITY: CONTAINS 1 DEATH DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U13616; AAA64834.1; -.  
 CC HSP; P55273; IBI8.  
 CC Genew; HGNC:494; ANK3.  
 CC MIM; 600465; -.  
 CC InterPro; IPR002110; ANK.  
 CC InterPro; IPR000488; Death.  
 CC InterPro; IPR000906; ZU5.  
 CC Pfam; PF00023; ank; 24.  
 CC Pfam; PF00531; death; 1.  
 CC Pfam; PF00791; ZU5; 1.  
 CC PRINTS; PR01415; ANKYRIN.  
 CC SMART; SM00248; ANK; 21.  
 CC SMART; SM00005; DEATH; 1.  
 CC SMART; SM00218; ZU5; 1.  
 CC PROSITE; PS50088; ANK\_REPEAT; 21.  
 CC PROSITE; PS50297; ANK\_REPEAT\_REGION; 1.  
 CC PROSITE; PS50017; DEATH\_DOMAIN; 1.  
 CC Cytokeleton; Alternative splicing; Repeat; ANK repeat.  
 FT REPEAT 73 102 ANK 1.  
 FT REPEAT 106 135 ANK 2.  
 FT REPEAT 139 168 ANK 3.  
 FT REPEAT 172 201 ANK 4.  
 FT REPEAT 203 230 ANK 5.  
 FT REPEAT 234 263 ANK 6.  
 FT REPEAT 267 296 ANK 7.  
 FT REPEAT 300 329 ANK 8.  
 FT REPEAT 333 362 ANK 9.  
 FT REPEAT 366 395 ANK 10.  
 FT REPEAT 399 428 ANK 11.  
 FT REPEAT 432 461 ANK 12.

3453 KPDRSFSQSLEVI-----EEGKVGPDDE-----KPSKSSSEKPKDKTQKSGA 3500

732 N-FWMSGDCGHFIWDRHTASHLMLEADNHNVCNLPHPDPILASSGIDYDIKIWSF 790

3501 QFTFLG-----RHPDRSVFPDITVFSYKVDPEEATPFTKV-ATKGLDFFD--PWSN 3547

791 LE-ESRIFNRKKLADEVITRNELMLEETRNITVP 823

3548 NRGDDEVFSQSKSREDETKPFLGLAVED-RSPATTP 3580

RESULT 38

TCGN2\_HUMAN STANDARD; PRT: 480 AA.

ID AC Q43493; Q43499; Q43500; Q15282; Q92760; Q43492; Q43501;

AD Q43493; Q43499; Q43500; Q15282; Q92760; Q43492; Q43501;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Trans-Golgi network integral membrane protein 2 precursor (Trans-

DE Golgi network protein TGN51) (TGN46) (TGN48) (TGN38 homolog).

OS TGNL2 OR TGN51 OR TGN46.

GN Homo sapiens (human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

[1]

SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND MUTAGENESIS OF TYR-430.

RP TISSUE=Liver, and Placenta;

RC MEDLINE=97063845; PubMed=9422759;

RX MEDLINE=97063845; PubMed=8907352;

RA Kain R., Angata K., Kerjasschi D., Fukuda M.;

RA Sumanabala S., Girotti M., Yaspo M.-L., Owen C.E., Perry A.C.,

RA Pananuma T., Nilsson T., Fried M., Banting G., Warren G.;

RT "Primate homologues of rat TGN38: primary structure, expression and functional implications";

RL J. Cell Sci. 109:675-685(1996).

RT FUNCTION: MAY BE INVOLVED IN REGULATING MEMBRANE TRAFFIC TO AND FROM TRANS-GOLGI NETWORK.

CC GOLGI NETWORK. CYCLES BETWEEN THE TRANS-GOLGI NETWORK AND THE CELL SURFACE RETURNING VIA ENDOSOMES.

CC ALTERNATIVE PRODUCTS: 3 ISOFORMS: TGN46, TGN48 AND TGN51 (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.

CC TISSUE SPECIFICITY: ISOFORM TGN46 IS WIDELY EXPRESSED. ISOFORM TGN51 IS MORE ABUNDANT IN FETAL LUNG AND KIDNEY. ISOFORM TGN48 IS BARELY EXPRESSED IN EMBRYONIC KIDNEY AND PROMYELOCYTIC CELLS.

CC THIS SWISS-PROT entry is copyright. It is produced through a collabor between the Swiss Institute of Bioinformatics and the EMBL outstat the European Bioinformatics Institute. There are no restrictions on use by non-profit institutions as long as its content is in no modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/annotation> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AF027516; AAC39542.1; -

DR EMBL; AF029316; AAB96908.1; -

DR EMBL; AF029313; AAB96908.1; JOINED.

DR EMBL; AF029314; AAB96908.1; JOINED.

DR EMBL; AF029315; AAB96908.1; JOINED.

DR EMBL; AF029316; AAB96906.1; -

DR EMBL; AF029315; AAB96906.1; JOINED.

DR EMBL; AF029313; AAB96906.1; JOINED.

DR EMBL; AF029314; AAB96906.1; JOINED.

DR EMBL; AF029315; AAB96906.1; JOINED.

DR EMBL; U62390; AAC39539.1; -

DR EMBL; AF027515; AAC39541.1; -

QY	438	SSPDSQRQSV	EAEGHHTHHQSDNNK	----	LSPKPQGTGEPVLSLH	YSTEGTTSTIKL	493								
Db	155	GAEPKTKORD	SPSKSGSEAQT	TKDVPNKSGADQ	TQPKGSS	----	KS 196								
QY	494	NFTDEWSSIA	SSRGISGHCKSE	QGBESFVQ	SSVQPEGSE	TKAPESSD	----	546							
Db	197	GAEDQTFD	VPNKSGAKQ	TQPKDGSN	KSGBAEQ	----	----	----							
QY	547	--VTYQEG	SAENPVNH	-----	INITQSDKFTAK	PLDLSNSGERNDLMD	RSCG	VP	597						
Db	255	EQPSRKDH	SPISNP	SDKNELPK	ADTNQ	LADKGLSP	HAFKTESGE	ETDL	----	304					
QY	598	ESASSEKAKE	PEFSDQ	TSTSA	NENNTPE	----	POFQTEATG	PSAAHETST	RD	SLQD	653				
Db	305	ISPPQEEV	KSESEPTD	VGPK	EA--	EDDDTG	PEEGSP	PEEK	EKMSG	SASSE--	NREGT	LS	D	361	
QY	654	TDDSDDD	VPVILP	GARYRAG	PGDRRS	AVARIQ	EFRRR	KERK	EMEELDT	LNI	RPLV	KWVY	713		
Db	362	STGSEK	DDL	-----	YPNGSG	-----	NGSAESSHFF	-----	AYLV	TAAI	----	LVA	V	L	400
QY	714	KHRNSRT	MIKEAN	FWGAN	FVMSG	737									
Db	401	IAHNKR	KRIIA	-----	FVLEG	416									

RESULT 39

BUD4\_YEAST

ID

BUD4\_YEAST

STANDARD;

PRT; 1447 AA.

AC

P47136;

DT

01-FEB-1996 (Rel. 33, Created)

DT

01-OCT-1996 (Rel. 34, Last sequence update)

DE

30-MAY-2000 (Rel. 39, Last annotation update)

DE

Bud site selection protein BUD4.

GN

BUD4 OR XJR092W OR J1905.

OS

Saccharomyces cerevisiae (Baker's yeast).

OC

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC

Saccharomycetales; Saccharomycetaceae; Saccharomyces.

OX

NCBI\_TaxID=4932;

[1]

RP

SEQUENCE FROM N.A.

RX

MEDLINE=9629555; PubMed=8707826;

RA

Sanders S.L., Herskowitz I.;

RT

"The BUD4 protein of yeast, required for axial budding, is localized

RT

to the mother/BUD neck in a cell cycle-dependent manner.";

RL

J. Cell Biol. 134:413-427(1996).

[2]

RP

SEQUENCE FROM N.A.

RX

Ramezani Rad M., Kirchath L., Hollenberg C.P.;

RA

Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.

RL

-1- FUNCTION: CO-ASSEMBLES WITH BUD3 AT BUD SITES. BUD4 AND BUD3 MAY

CC

COOPERATE TO RECOGNIZE A SPATIAL LANDMARK (THE NECK FILAMENTS)

CC

DURING MITOSIS AND THEY SUBSEQUENTLY BECOME A LANDMARK FOR

CC

ESTABLISHING THE AXIAL BUDDING PATTERN IN G1.

CC

-1- SIMILARITY: CONTAINS 1 PH DOMAIN.

CC

-1- CAUTION: REP 2 SEQUENCE DIFFERS FROM THAT SHOWN IN THE N-TERMINAL

CC

SECTION DUE TO FRAMESHIFTS.

CC

This SWISS-PROT entry is copyright. It is produced through a collaborat

CC

between the Swiss Institute of Bioinformatics and the EMBL outstation

CC

the European Bioinformatics Institute. There are no restrictions on

CC

use by non-profit institutions as long as its content is in no w

CC

modified and this statement is not removed. Usage by and for commerc

CC

entities requires a license agreement (See <http://www.isb-sib.ch/announcement>

CC

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR

EMBL; U41641; AAB17116.1;

DR

EMBL; Z49592; CAA89620.1; ALT\_FRAME.

DR

EMBL; Z49591; CAA89619.1;

DR

\_SGD; S0003852; BUD4

DR

InterPro: IPR001849; PH.

DR

Pfam: PF00169; PH; 1.

DR

SMART; SM00233; PH; 1.

DR

PROSITE; PS50003; PH\_DOMAIN: 1.

RESULT 40  
YAU9\_SCHPO STANDARD; PRT; 1275 AA.  
ID YAU9\_SCHPO  
AC Q10164;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein C26A3.09c in chromosome I.  
DE SPAC26A3.09C.  
GN Schizosaccharomyces pombe (Fission yeast).  
OS Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=972;  
RC MEDLINE=21848401; PubMed=11859360;  
RX WOOD V., William R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
RA Sgouros J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Feltwell J., Fraser A.,  
RA Gentles S., Goble A., Hamlin N., Harris D., Hiddleston J., Hodgson G.,  
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
RA Welljens J., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Gaillardin C., Tallaie V.A., Garzon A., Thode G.,  
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
RT "The genome sequence of Schizosaccharomyces pombe.";  
RL Nature 415:871-880(2002).  
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 RHO-GAP DOMAIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Z69240; CAA93232.1; -;  
DR HSPG; Q07960; IRGP.  
DR InterPro; IPR001849; PH.  
DR InterPro; IPR00198; RhoGAP.  
DR Pfam; PF00169; PH; 1.  
DR Pfam; PF00620; RhoGAP; 1.  
DR SMART; SM00233; PH; 1.  
DR SMART; SM00324; RhoGAP; 1.  
DR PROSITE; PS50003; PH\_DOMAIN; 1.  
KW Hypothetical protein.  
FT DOMAIN 719 836 PH.  
SQ SEQUENCE 1275 AA; 143586 MW; 136DAA26BA41BBE CRC64;  
Query Match 3.5%; Score 159; DB 1; Length 1275;  
Best Local Similarity 19.3%; Pred. No. 0.71;  
Matches 106; Conservative 79; Mismatches 198; Indels 166; Gaps 22;



```

QY 237 TTGMVARFTPSHLNKNKSCRVTSICYSYSDYIYLFDPKDDTARELKTPSAE 296
Db 9 TTGFFLRSEPT-----TCRITTFAYSAE-----IIYFHAETTFEFKIMEGNELEKYSKI 59
QY 297 ERREELRQPPVKRLRLRGDWSGTGPRARPESEREDRGEOSPNNVLMQ-----RMSDMLSRWF 353
Db 60 ELNSEL-----WED-----EEEDNSGIVSQNERLMLKLVNKQREIYYELH 99
QY 354 EASVAQSNRGRGRPRGGTSDISTLPTVPSPDLEVSSETAMEVDTPAEQFLQPS 413
Db 100 KDLKVKKNDTSLRMLSKYESTDFFPSQPSRANSPOSDSYSSPYE-----KGLPKISL 156
QY 414 SSTMSAQAHSTSSPTSPHSTPLSSPDSRQSVASGHHHQ-----SDN 461
Db 157 KS-----SKDVPTASAH-----ISSDHEKSSSVSLSALNNYKTTDIKARSLDLRLSDM 205
QY 462 NNEKL-----SPKPGTGPEVLS-----LHYSTEGT----- 486
Db 206 TRPKLLNLTKRSHRSSEPCGASSPVTSPILKDSOKERIQALRNKAITYSVSTESAERID 265
QY 487 -----TSTIKLFTDEWSSIASRRGIGSHCKSEGOEESF-----VPOSSV 528
Db 266 SIRDNLSPISLATSSFRPITKTPFNDSNISIIDPKDNNKQDHFAEIEDELROOFL 325
QY 529 QPPEGDSKAPESSESDVTKYQGVSAENPVENHINITOSDKF----- 572
Db 326 DIKVRANASSPRKKSISIVK-PHGISSPKHSTNNLS-SKSGKFHSDFRVVSENVLLQAR 383
QY 573 --TAKPLDSNGERNLNLDRSCGVPESASSEKAKPETSDOTSTESATNENNTNPEQ 630
Db 384 SETNSPIENK-EANNF-LAPTSNVPAYSTPARPESP-----PPP 423
QY 631 FQTEATGPSAHEETSTRDSALOTDDDDPVLIPGARYAGPGD-RRSAVARIQE----- 685
Db 424 ISSSSTTPRPDDKPSLPPLGLEDNDS-----LSLQKTGSSDTRRSFSTLKIPDSD 475
QY 686 --FFRRRKE 692
Db 476 ICFTRRRSD 484

```

Search completed: March 18, 2003, 15:35:01  
Job time : 64 secs

10

10

GenCore version 5.1.4\_p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 18, 2003, 15:32:49 ; Search time 23 Seconds  
(without alignments)  
3594.591 Million cell updates/sec

Title: US-09-781-693A-2

Perfect score: 4527

Sequence: 1 MSRGSGYPHLLWDVRSLSG.....RLEGDRSGSGQENENEDEE 860

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: PIR\_73:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	590.5	13.0	754	2 T50803	hypothetical prote
2	547.5	12.1	809	2 T40574	quanine nucleotide
3	448.5	9.9	481	2 T47502	hypothetical prote
4	429.5	9.5	747	2 S37694	gene PC326 protei
5	405.5	9.0	493	2 T04951	hypothetical prote
6	384.5	8.5	426	2 T05676	hypothetical prote
7	220.5	4.9	489	2 A45988	dentin matrix acid
8	204.5	4.5	1217	2 T00270	hypothetical prote
9	202	4.5	5170	2 T15348	hypothetical prote
10	199	4.4	550	2 T29919	hypothetical prote
11	198.5	4.4	406	2 S38170	SRP40 protein - ye
12	198	4.4	3507	2 T34513	hypothetical prote
13	193	4.3	852	2 T06310	hypothetical prote
14	190.5	4.2	1367	1 S48478	glucan 1,4-alpha-g
15	188	4.2	1702	2 A41859	IgA-specific metal
16	187.5	4.1	2004	2 F95133	immunoglobulin A1
17	186	4.1	1694	2 H64106	IgA-specific metal
18	183	4.0	727	2 T29612	IgA-specific metal
19	183	4.0	2271	2 F90073	hypothetical prote
20	180.5	4.0	868	2 F89500	hypothetical prote
21	178	3.9	1027	2 T46481	hypothetical prote
22	178	3.9	1051	2 T48933	hypothetical prote
23	177.5	3.9	1063	2 D86731	protein T27A10.6 [
24	177.5	3.9	1658	2 S55101	WD repeat domain p
25	176.5	3.9	734	2 B42680	hypothetical prote
26	176.5	3.9	802	2 T21315	nucleolus-cytoplas
27	176.5	3.9	5327	2 T13564	hypothetical prote
28	176	3.9	1459	2 T32271	microtubule-associ
29	175.5	3.9	1849	2 C41859	IgA-specific metal

30	175	3.9	534	2 T39903	serine-rich protei
31	175	3.9	1063	2 T38732	probable helicase
32	175	3.9	1829	2 T24583	hypothetical prote
33	174.5	3.9	787	2 A48819	nuclear autoantige
34	174.5	3.9	1587	2 G86467	hypothetical prote
35	174	3.8	1200	2 A46194	neurofilament prot
36	174	3.8	2232	2 T34434	hypothetical prote
37	173	3.8	895	2 T45738	hypothetical prote
38	173	3.8	1641	2 T38614	helicase II - huma
39	173	3.8	1963	2 B98002	IgA-specific metal
40	172.5	3.8	1233	2 T30989	serine/threonine p
41	172	3.8	664	2 S60062	hevin precursor -
42	172	3.8	1007	2 T01437	hypothetical prote
43	171.5	3.8	1337	2 T30291	dextranase - Strep
44	170.5	3.8	1189	2 S56852	hypothetical prote
45	169.5	3.7	619	2 JX0052	apopolysialoglycop
46	169.5	3.7	1092	1 JN0635	neural cell adhesi
47	169.5	3.7	1589	2 C44766	defective chorion-
48	168	3.7	910	1 S73361	dnaj homolog prote
49	167.5	3.7	2342	2 T13412	hypothetical prote
50	166.5	3.7	1359	2 T34036	hypothetical prote
51	166.5	3.7	2722	2 T20532	hypothetical prote
52	166	3.7	643	2 T19199	hypothetical prote
53	165.5	3.7	891	2 G84693	probable proline-r
54	165.5	3.7	1210	2 T39410	AF-4 protein, spli
55	165.5	3.7	1435	2 A37793	erythrocyte-bindin
56	165.5	3.7	3924	2 S37431	ankyrin 2, neurona
57	164.5	3.6	660	2 JMW067	chitinase (EC 3.2.
58	164.5	3.6	1419	2 T30531	agglutinin-like ad
59	164	3.6	1188	2 T05324	hypothetical prote
60	163.5	3.6	971	2 T19431	hypothetical prote
61	163.5	3.6	1747	2 AC1842	WD-40 repeat prote
62	163	3.6	1466	2 A36426	SPA2 protein - yea
63	163	3.6	6842	2 T29757	protein UNC-89 - C
64	162.5	3.6	699	2 T38073	nucleolar phosphop
65	162.5	3.6	1213	2 A58198	serine/proline-ric
66	162.5	3.6	1630	2 A53577	ascites sialoglyco
67	162	3.6	727	2 T24284	hypothetical prote
68	162	3.6	1032	2 T34433	hypothetical prote
69	161.5	3.6	671	2 S53407	CHS5 protein - yea
70	161	3.6	1390	2 T44004	trfa protein - sli
71	161	3.6	3938	2 T42761	Bassoon protein -
72	160.5	3.5	1072	2 A86827	hypothetical prote
73	160.5	3.5	1175	2 T25634	hypothetical prote
74	160.5	3.5	1356	2 T18521	beta transducin-li
75	160.5	3.5	4377	2 A55575	ankyrin 3, long sp
76	159.5	3.5	1386	2 T49316	profilaggrin relat
77	159.5	3.5	1560	2 T30282	calcium-binding pr
78	159.5	3.5	2481	2 D90011	FmtB protein (limp
79	159	3.5	961	2 E86245	hypothetical prote
80	159	3.5	1275	2 T38397	probable GTPase ac
81	159	3.5	1310	2 T40135	probable involveme
82	159	3.5	1711	2 AD1842	WD-40 repeat prote
83	158.5	3.5	1104	2 S53310	probable membrane
84	158.5	3.5	2738	2 E88320	protein F07A11.6 [
85	158	3.5	1280	2 T00365	hypothetical prote
86	158	3.5	2774	2 A43359	microtubule-associ
87	157.5	3.5	680	2 A43800	nuclear autoantige
88	157.5	3.5	1912	2 T29088	vitellogenin I pre
89	157	3.5	1151	2 T33777	hypothetical prote
90	157	3.5	1286	2 A88396	protein M01E10.2 [
91	157	3.5	1613	2 JC6510	ras-responsive ele
92	156.5	3.5	532	2 T06029	hypothetical prote
93	156.5	3.5	713	2 JN0133	WD-40 repeat regul
94	156.5	3.5	3942	2 T42730	Bassoon protein -
95	156	3.4	1015	2 S68141	nuclear protein HI
96	156	3.4	1211	2 T42230	AF4 protein - mous
97	156	3.4	1791	2 T02345	hypothetical prote
98	156	3.4	2957	2 T33152	hypothetical prote
99	155.5	3.4	505	2 B64560	poly E-rich protei
100	155.5	3.4	786	2 T16509	hypothetical prote
101	155.5	3.4	1020	1 QFHDH	neurofilament trip
102	155.5	3.4	1217	2 T42625	AF-4 protein - mou

SUM1 protein - yea  
hypothetical serin  
hypothetical prote  
nuclear histone-bi  
neurabin - rat  
gravin - human  
hypothetical prote  
transcription fact  
cell proliferation  
tupl-like enhancer  
hypothetical prote  
neurofilament prot  
hypothetical prote  
a-agglutinin core  
secreted acide phos  
probable serine/th  
hypothetical prote  
neurofilament trip

103 155 3.4 1062 2 S61196  
104 155 3.4 1131 2 T41144  
105 155 3.4 1229 2 T25697  
106 154.5 3.4 1290 2 A25680  
107 154.5 3.4 1095 2 T43275  
108 154.5 3.4 1684 2 JW0057  
109 154.5 3.4 2218 2 B84683  
110 154 3.4 1093 2 A47212  
111 154 3.4 2938 2 T30249  
112 153.5 3.4 519 2 S45345  
113 153.5 3.4 831 2 T48442  
114 153.5 3.4 913 2 T24485  
115 153 3.4 699 2 T01029  
116 153 3.4 725 2 A41258  
117 153 3.4 888 2 T46726  
118 153 3.4 1192 2 T18611  
119 153 3.4 2526 2 T20531  
120 152.5 3.4 606 2 A43427

ALIGNMENTS

hypothetical protein T30N20\_210 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 21-Jul-2000  
C:Accession: T50803  
R:Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Bancroft, I.; Meyer, M.; ...  
Submitted to the Protein Sequence Database, July 2000  
A:Reference number: Z25240  
A:Accession: T50803  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-754 <BEV>  
A:Cross-references: EMBL:AL365234  
A:Experimental source: cultivar Columbia; BAC clone T30N20  
A:Note: T30N20\_210

Query Match 13.0%; Score 590.5; DB 2; Length 754;  
Best Local Similarity 24.6%; Pred. No. 2.9e-24;  
Matches 207; Conservative 114; Mismatches 279; Indels 245; Gaps 28;

Db 343 -----TTPQNGFHR--SNAATVKKCTELVEIAKWSLE 373  
QY 401 VDT-----PAEQFLQPTSTMSAQAHSTSSP-----TESPHSTPLSPDSE 443  
Db 374 EGTDFVYAEAAANEVLDHNSDIESALRECLCTRAALLKKRWKNDH-----MAVRDCH 429  
QY 444 QRSVEASGHHTHQSDDNNNEKLSKPGTGEVPLSLHYSTEGTTSTIKLNFDEWSSIA 503  
Db 430 NARRIDASFKAHYIMSEALQQL-----GKCEALDFATAAHMNPSDADIYAKVESIK 483  
QY 504 SSSRGIGSCKEGQESFVQSSVQPPGDSSTK-----APEESSEDVTYKQ 551  
Db 484 RDLQAAG-----AEKNEET-----GAGTRVLISLSDILYRSEANSSSHDMSRSE 538  
QY 552 EGVSAENPVENHINTQSKFTAKPLDLSNGSRNDLNRSCGVPPEESASSKAKEPET 610  
Db 529 REDSDYDEELELDIQISLSD--EGRTDSDNS--MRGSLNL-----RIHYGDDKPME--NT 579  
QY 611 SDQTSFESATNENNTNPEQFQTEATGPSAHEETSTRDSALQDTDDSDDDPVLIPGARYR 670  
Db 580 VDNASSGTASSQN-----DRTSYOPEGAID----- 605  
QY 671 AGPDRSARVAIQEFRRRKRERKEMBELDTLNRPLRVKMYKGRHSRTMKEANFWG 730  
Db 606 -----MKRRYVGHCVNGTDLKQASFLG 627  
QY 731 --ANFVMSGDCGHFTWDRHTAEHMLLEADNIVNCLQHPDPPILAGSDIDYDIKI 788  
Db 628 QRGEYIASGSDGGRWFIWEKOTGRIMKLVGDESVLNOCIOCHPFDVSVVATSGIDNTIKI 687  
QY 789 SP 790  
Db 688 SP 689

RESULT 2  
T40574  
guanine nucleotide binding protein beta subunit-like - fission yeast (Schizosaccharom  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C:Accession: T40574  
R:Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, January 1999  
A:Reference number: Z21938  
A:Accession: T40574  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-809 <SEE>  
A:Cross-references: EMBL:AL035226; PIDN:CAA22832.1; GSPDB:GN00067  
A:Experimental source: strain 972h-; cosmid c609  
C:Genetics:  
A:Gene: SPAC609.03  
A:Map position: 2  
A:Introns: 21/1; 52/2; 280/3; 780/2

Query Match 12.1%; Score 547.5; DB 2; Length 809;  
Best Local Similarity 24.8%; Pred. No. 6.4e-22;  
Matches 236; Conservative 124; Mismatches 308; Indels 285; Gaps 41;

QY 22 BDPS-RLRSRYLGRREFTORLKEATLVNHDGCVNTICWNTGEYILSGSDTKLVISNP 80  
Db 20 QDPSHEVDQRMQFHSLVRLSQEQLEGHQCVNALAMNSGSLISGSDDLRLINWY 79  
QY 81 YSRKVLTIIRSGHRANIFSAKFLPCTNDKQIVSCGSDGVI--FYTN-----VEODAETNR 133  
Db 80 SSRKLLHSIDTGTANIFCTKVPETSDLVVSGAGDAEVRFLNFTSLSGRAEDDNIAP 139  
QY 134 QCQFTCHYGTTEIMTPNDPFTYFSLCGEDGVRVWFDPRIKTSVSC-----TKEDCKDDILL 188  
Db 140 SALLYQCHTRRVKKLAVPCGNPNVWVSASEDGLRQHDRESTSCPPAGTAHQECR--SVLL 198  
QY 189 NCRRAATSVACPP-----IPYVLAVGSDSSSVRIYDRMLGTRATG----- 230  
Db 199 DLRSAGKALADPPKOTLSLKSDISATRPHLLLVGGSDAFARLYDRMLPPLASSRKM 258  
QY 231 -----NYAGRTGTGMVAFIPSHLNNK---SCRVTSLCYSDGQELVSYSDIYLF 281  
Db 259 PPPPCVNY-----FCPMHLSRGRTNLHLTHVTSPNGEELVSYSGHERVLAN 307  
QY 282 PKDDTARELKTSAEERERELPPVKRLRLRGDWDSDTGPRAPEERERDQSPNVSL 341  
Db 308 VNGTGTIMQYTPGDVNDLFSF-----SNLHDSVSPQVS- 342  
QY 342 MQRMSDMLSRWFEASEVAQNGRGRSRPRGGTQSOSDITLPTVPSPDL--EVSETAME 400

QY 27 LRSRYLGRREFTOR-----LKLKATLVNHDGCVNTICWNTGEYILSGSD 71  
Db 6 LSLRLQDYRDWFQKTSRDIYGNSTWLTGIDLQKELTGTGCVNTLDWSADGEFLSGSD 65  
QY 72 DTKLV---ISNPNYSRKVLTIIRSGHRANIFSAKFLPCTNDKQIVSCGSDGVIYFTVQED 128  
Db 66 DTRLIWVDVFNKYKPRHL--ISTGHVQNIFFSAKFPVYNNRQILSASGDKLIKLFOLDSS 123  
QY 129 -----AFTNQCOFTCHYGTTEIMTPNDPFTYFSLCGEDGTVWFDTIKTSCTKE 180  
Db 124 KEGMDHGMETQTRC--WSCALDSVKNIYVPCDNG--HTFLVCSGDTARQYDIRPHVCNQD 181  
QY 181 -DCKDDILLIN--CRRRAATSVACPPPIYVLAVGSDSSSVRIYDRMLGTRATG-----TGNY 232

Matches 167; Conservative 72; Mismatches 194; Indels 437; Gaps 16;	
QY	13 DVKRKSLGLEDPRLRSRYLGRREFIQRULEATLNVHDGCVNTICWNTGEYILSGSD 72
Db	17 E1FNREIGFSHPITISRRISASEGRVKKLDLYGKLNHGGCVNAVEFNSTGDLVSGSD 76
QY	73 TKLILSNPYSRKVLITISGHRANISAKFLPCTNDKQIVSGDGVIYFNVQD--AE 130
Db	77 RQIMLWNLWSGRKLSYPSGHCFYQFKFIPFTDDRTIITSGADQVRLGQILLENKVE 136
QY	131 TNRCQFTCHYTTEIMTPNDPFTLSCGDEGVNMFEDRIKTSCKEDCKDILINC 190
Db	137 TKRLGR--HNGRVYKLAIVLPCDPNVFSCGDEGFVQHFDIR-SNSATMVLVLYSPFTQC 192
QY	191 RE-----AATVAICPPYIYLAVGSDSSVRIYDRMLGLTRATGYAGRGTTGMVAR 243
Db	193 RKHSSIRLNSIAIDPRNSYIYLAVGSDSEYARVYDITRRVOLAPVCRHV--LPDAPVNT 250
QY	244 FIPSHL-NNKSQRTVSLCYSDGQFELVSYSDVLYLFDKDDTARELKTPSAERREEL 302
Db	251 FCPHRLRETNVHITGLAYSAG-ELLVSYNDELIYLF-----287
QY	303 RQPPYKRLRLRGDSWDTGTPRAPESERDERGEQSPNVSLMQMSDMLSRWFEEASEVAQS 362
Db	288 -----EK 289
QY	363 NRGGRSRPRGGTSQSDISTLPTVPSPDLEVSETAMEVDTPAEQFLQPSSTMSAQAH 422
Db	290 NMWYG-----294
QY	423 STSPSTSPHSTPLLSSPDSEQROSVESAGHHTHQSDNNNEKLSPKPGTGEVLSLHYS 482
Db	295 --SSPV-----298
QY	483 TEGTTTSTIKLNTDEWSSSIASSRGISGCKSEGESEFVPOSSVQPPGDSSETKAPEE 542
Db	299 -----298
QY	543 SSEDVTKYQGVSAENPVENHINITOSDKFTAKPLDSNGERNDLNRSCGVPESASS 602
Db	299 -----SVSP 302
QY	603 EKAKEPETSQDTESATNENNTNPEPQTEATGQSAHEETSTRDSALQDITDSDSDPV 662
Db	303 EKQLEME-----EPQ-----312
QY	663 LIPGARYRAGPGRDRSARAVARIOEFFRRRKRKEMEELDTLNIRRPVLMVYKYGHRNRTM 722
Db	313 -----VYIGHNAQT- 322
QY	723 IKEANFWGAN--FVMSGSCGHIPIWDRHTAHLMLLEADNVNVNCLQPHDPDPTILASSG 780
Db	323 VRGVNFFGPNDEYVTVSGSDCGHIFWKKGKGLVRAMGVDRRVVQLESHPIPLASC 382
QY	781 IDYDIKWIPLESRIENFKLADEVITNELMLEETRTNITVPASFM--LRMLASLNHR 838
Db	383 IEKSVKLWTPMNSDNLVSLPEKIDKVELNVRGREQDSRVTLTPDVIHVLRLQRQTSAF 442
QY	839 ADRLGDRSEGGQENE-----NEDE 859
Db	443 TERRYVSTDIGSDGNDARFIASLVANDDE 472
RESULT 4	
S37694	
gene PC326 protein - mouse	
C:Species: Mus musculus (house mouse)	
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 26-May-2000	
C:Accession: S37694; S27875	
R:Berghsagel, P.L.; Timblin, C.R.; Eckhardt, L.; Laskov, R.; Kuehl, W.M.	
OncoGene 7, 2059-2064, 1992	
A:Title: Sequence and expression of a murine cDNA encoding PC326, a novel gene ex	
A:Reference number: S37694; MUID:93026383; PMID:1408147	

A:Accession: S37694  
A:Molecule type: mRNA  
A:Residues: 1-747 <ECK>  
A:Cross-references: EMBL:M95564; NID:g200240; PIDN:AAA39895.1; PID:g200241  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 13-Aug-1999  
C:Accession: T04961; T05784  
R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Hoheisel, J.; Mewes, H.W.; Mayer, K.  
submitted to the Protein Sequence Database, February 1999  
A:Reference number: Z15392  
A:Accession: T04961  
A:Molecule type: DNA  
A:Residues: 1-493 <BEV>  
A:Cross-references: EMBL:AL035522  
A:Experimental source: cultivar Columbia; BAC clone T12J5  
R:Bevan, M.; Purnelle, B.; Boutry, M.; Goffeau, A.; Hoheisel, J.; Mewes, H.W.; Mayer, K.  
submitted to the Protein Sequence Database, April 1998  
A:Reference number: Z15451  
A:Accession: T05784  
A:Molecule type: DNA  
A:Residues: 398-493 <BEW>  
A:Cross-references: EMBL:AL022023  
A:Experimental source: cultivar Columbia; BAC clone M4E13  
C:Genetics:  
A:Map position: 4  
A:Introns: 43/3; 127/3; 371/3  
A:Note: T12J5.10; M4E13.190

Query Match 9.0%; Score 405.5; DB 2; Length 747;  
Best Local Similarity 18.7%; Pred. No. 1.2e-15;  
Matches 164; Conservative 76; Mismatches 173; Indels 465; Gaps 18;

QY 8 PHLLMDV-----RRRSGLGLEDPSRLSRYLGRREFIORLKLEATLNVDGCVNTICNDG 63  
DB 308 PRPRNVLNLRDLQLG--SSGRFVEACGARLFQVRSLEHVFEGHSGCVNTVHFNQHG 365  
QY 64 EYILSGDDTKLVISNPYSRKVLTTRSGHRANIFSAKFLPCTNDRQIVSCSDGQVIFYT 123  
DB 366 TLLASGSDDLKVIWDMWKRSVNFDSGHKNLLQAKFLPNCNDAILAMCGRDQVRVA 425  
QY 124 NVQDAETNRCQOFTCHYGTTEIMVNDPPTFLSCGEGDTVRWFDTRIKTSCTREDCK 183  
DB 426 QLSAVAGTHMTKLVKHGGASHRLGLEPDSFPRFLTSGEDAVVFNIDLRQAHFASK 481  
QY 184 DDILI-----NCRRAATVAICPPPIPYLVAVGCDSSSVRIYDRMLGTRATGNTAGRTTG 239  
DB 482 --LLVIKDGDKKVLGYTVFVNPNVYQFVGGQDQFRIYDQKIDENV-----NNG 531  
QY 240 MVARFIPSHL--NNKSCRVTSLCYSEDGQELVSYSDIYLFDPKDDTARELKTPTSAAE 297  
DB 532 VLKFCPPHLLSSDPYAHITSLMYSYDGTETILASYNEDIYF----- 574  
QY 298 RRELOPPVKRLRLGDSWDTGTPRAPERERDGEQSPNVSLMQMSDMLSRWFEEAS 357  
DB 575 ----- 574  
QY 358 EVAQSNRGRSRPRGTSQSDISTLPTVPSPDLEVSETAMEVDTPAEQFLQSTSTSM 417  
DB 575 ----- 574  
QY 418 SAQAHSTSPSPESHSTPLLSPPDSEQSQVEASGHHHTHQSDNNNEKLSPKPGTGPVL 477  
DB 575 ----- 574  
QY 478 SIHYSTEGTTTSTIKLFTDEWSSIASSSRGIGSHCKSEGESEFVQSSVOPPEGDSET 537  
DB 575 ----- 574  
QY 538 KAPEESEDVTKYQGVSAENPVENHINITQSDKFTAKPLDSNGSERNLNLDRSCGVPE 597  
DB 575 ----- 574  
QY 598 ESASSEKAKEPETSQSTESATNENNTNPEQFQTEATGPSAHEETSTRDSALQDSTD 657  
DB 575 -----NSSDS 579  
QY 658 DDDPVLIPGARYRAGDRRSARARIQEFRRRKRKEMEELDTLIRPLVKNYKGRH 717  
DB 580 D-----GAQY-----AKR-----YKGRH 592  
QY 718 NSRTMKEANFWG--ANFVMSGDCGHIFIDWRHTAHLMLLEAD--NHVVNCLQHPFDP 774  
DB 593 NNST--VKGYVFGPREFVMSGDCGHIFIEKSSCOIQVLEADGGTINCIDSHPLP 651  
QY 775 ILASSGIDYDIKIPLESRIFNKLA--DEVITRNELMLEE--TRNITVTPAS---FM 827  
DB 652 VLASSGLDHEVKIWSPIAEP---SKLAGLKNVINKIKLRDNFTLRHTSLFNNSMLCFL 708  
QY 828 LRLASLNH-----IADRLGDRSGSGQENENEDE 859  
DB 709 MSHVTSQNTGRSWRGIRINAGGGDFSDSSSSSEETNQE 746

RESULT 5  
T04961  
hypothetical protein T12J5.10 - Arabidopsis thaliana  
N:Alternate names: hypothetical protein M4E13.190  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 13-Aug-1999  
C:Accession: T04961; T05784  
R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Hoheisel, J.; Mewes, H.W.; Mayer, K.  
submitted to the Protein Sequence Database, February 1999  
A:Reference number: Z15392  
A:Accession: T04961  
A:Molecule type: DNA  
A:Residues: 1-493 <BEV>  
A:Cross-references: EMBL:AL035522  
A:Experimental source: cultivar Columbia; BAC clone T12J5  
R:Bevan, M.; Purnelle, B.; Boutry, M.; Goffeau, A.; Hoheisel, J.; Mewes, H.W.; Mayer, K.  
submitted to the Protein Sequence Database, April 1998  
A:Reference number: Z15451  
A:Accession: T05784  
A:Molecule type: DNA  
A:Residues: 398-493 <BEW>  
A:Cross-references: EMBL:AL022023  
A:Experimental source: cultivar Columbia; BAC clone M4E13  
C:Genetics:  
A:Map position: 4  
A:Introns: 43/3; 127/3; 371/3  
A:Note: T12J5.10; M4E13.190

Query Match 9.0%; Score 405.5; DB 2; Length 493;  
Best Local Similarity 18.9%; Pred. No. 1.4e-14;  
Matches 162; Conservative 74; Mismatches 151; Indels 471; Gaps 20;

QY 9 HLLMDVRRKRSGLGLEDPSRLSRYLGRREFIORLKLEATLNVDGCVNTICNDGTEYILS 68  
DB 17 HPVYDFWRREVGISRRNFSNRSASENLVLRLEIYKLEKHKGCVNTVFNAGDVLIS 76  
QY 69 GSDTKLVISNPYSRKVLTTRSGHRANIFSAKFLPCTNDRQIVSCSDGQVIFYTNVED 128  
DB 77 GSDRRVVLWDLQNLGNVLSFHSGHANNVFOAKFMPFSDDRITVTCADGMF---DLRTE 133  
QY 129 AETNRCQOFTCHYGTTEIMVNDPPTFLSCGEGDTVRWFDTRIKTSCTREDCKDDILI 188  
DB 134 APTE---LFTC-----RSVDPRRN-----MDATQL 156  
QY 189 NCRRAATVAICPPPIPYLVAVGCDSSSVRIYDRMLGTRATGNTAGRTGMVA---RPI 245  
DB 157 N-----ALADPRNSNLFNFGVGMEEYARLYDIR-----RFQEGGLNGFTRAADHFC 202  
QY 246 PSHL--NNKSCRVTSLCYSEDGQELVSYSDIYLFDPKDDTARELKTPTSAAERRELQ 304  
DB 203 PPHLLIGNEDVGITGLAFSEQ--SELLVSYNDEFIYLFTP----- 239  
QY 305 PVKRLRLGDSWDTGTPRAPERERDGEQSPNVSLMQMSDMLSRWFEEASEVAQSNR 364  
DB 240 ----- 239  
QY 365 GRGRSRPRGTSQSDISTLPTVPSPDLEVSETAMEVDTPAEQFLQSTSTMSAQAHST 424  
DB 240 GMGLG----- 244  
QY 425 SSPTESPHSTPLLSPPDSEQSQVEASGHHHTHQSDNNNEKLSPKPGTGPVLSLHYSTE 484  
DB 245 -----SNIPSPSPIS----- 254  
QY 485 GTTSTIKLNFDEWSSIASSSRGIGSHCKSEGESEFVQSSVQPEGSSEKAPES 544  
DB 255 -----KSPVSKS 261  
QY 545 EDVTKYQGVSAENPVENHINITQSDKFTAKPLDSNGSERNLNLDRSCGVPEASSEK 604  
DB 262 -----ESSSPK 268



Db 394 EDSOQGLQSQASRESRSQESQSEDSRSENRSDD 431

RESULT 8

T00270

hypothetical protein KIAA0596 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 21-Jul-2000

C:Accession: T00270

R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.

DNA Res. 5, 31-39, 1998

A:Title: Prediction of the coding sequences of unidentified human genes. IX. The complete

A:Reference number: Z14086; PMID:98290545; PMID:9628581

A:Accession: T00270

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1217 <NAG>

A:Cross-references: EMBL:AB011168; NID:g3043715; PIDN:BAA2522.1; PID:g3043716

A:Experimental source: brain

A:Genetics:

A:Note: KIAA0596

Query Match 4.5%; Score 204.5; DB 2; Length 1217;

Best Local Similarity 20.6%; Pred. No. 0.0027;

Matches 187; Conservative 121; Mismatches 351; Indels 247; Gaps 43;

QY 39 QRLKLEATLVHDCVNTICWN--DTGEYILSGSDTKLV---ISNPYSKVLTTIRSG 92

Db 218 EMLKEA---HDEILGLEYKPTGTLKLLASRRDLIHLVDAGREYS---LQOTLDE 270

QY 93 HRANIFSAKFLPCTNDKQIVSCSGDGVIFYTNVEDQAEETNRQCOFT-CHY-----GTTVEI 147

Db 271 HSSITAVKFAASDQGRVMISGADKSIYFRTAOKSGD---GVQFTRHHVVRKTKLYDM 327

QY 148 MTPVNDPTFLSCGDEGVVRFNFDRIKTSCTKEDCK-----DDILNCRRAATSAVACPP 202

Db 328 DVEPSWKYTAIGC-QDRNIRIFN--ISSGOKKLFKSGDEGTLIKVQ-----TDP 376

QY 203 IPYLLAVGCSDDSVRIYDRMLGTRATGNAGRGVYTGVMARFIPSHLNKSCRVTSIYCS 262

Db 377 SGIIYATSCDKNLSIFD-----FSSGECVATMFGH-----SEIVTGMKFS 417

QY 263 EDQBEILVSYSSD-YIYLPDKDDTARELKTPTSAEREE---LRQPPVKRLRLRG--- 314

Db 418 NDCKH-LISVSGDSCIFVRLSEMSTIMRQRLAELRQGRKGQGGQSSPORASGPNRH 476

QY 315 ---DWSDTGPRAPRESERDGG---EQPNVSLMQR-----MSDMLSRWFE 354

Db 477 QAPSMSPGALSDSDSDEGEDETEELPALPVLAKSTKALASVPSPALPRSLHW--- 534

QY 355 EASEVAQS-----NRRGRSRPRG-----GTSQSD 379

Db 535 EMSRAQSVGLDPAPAAANPGRRRGRVQVQVGLSVRSMGLDLRLQETLAPSLQDPSQDS 594

QY 380 ISTLPVSPDLEVSETAMEVDTPAEQFLQPTSTSTMSAQAHSTSPSPHSTPLLS 439

Db 595 LATIPSGPKRHQGEALETSL-----TSQNEKPPRQAQPCSYPIIRLLSQ 641

QY 440 PDSEQRQSVAS---GHHTTHQSDN-----NNEKLSPK 469

Db 642 EGVFAQDLPAFIEDGIYVPEPDNPTMDTSEFQVQAPARGTLGRVYPGSRSEKHP- 700

QY 470 PGVGEVLSLHYTEGTTSTIKLNTDEWSSIASRSGISHCK--SEGOESESFVQSS 527

Db 701 ----DSACSVDYSS--SCUSPSHPHTDESSTPLSDVGLSSDLEPAEDDEEEEGG 754

QY 528 VQP---PEGDSETKAPE--SEEDVTKYQGVSAENPV---ENHINTQSKF-----TAKP 576

Db 755 MGYGLQEGSPQIPDQQLKHFETLASGAAPGVQVPERSESRISIRSLFIQVQTRP 814

QY 577 LDNSGGRNDLNL--DRSCGVPEESASSEK---AKEPETSDQTSFESATNNTNPEQ-- 630

Db 815 LREPSPSSSLALMSRPAQVQASGEQPRNGANPPGAPPEVEPSSG-----NPSFQA 868

QY 631 -----FQTEATGFSAAHEETSTRDSALQDSTDSDDDPVLIPGARYA-----GGDR 676

Db 869 ASVLLPCRNLNPDSSWAPKRVATASPFSGIQAQSVHS---LVPQERHEASLOAPSPGAL 925

QY 677 RSAVARIQEFFFFRRKKEKMEELDTLNIRPLVMYVKGH-----RNSRTMKKEANFWG 730

Db 926 LS-----RETEAQDGLGSLPPADGPPSRPHSYQNPTTSSMAKISRISVSG 970

QY 731 ANFVMSGSDCGHIFINDRHTAEHLMLLEADNH--VVCNQPHPPDPILLASSGIDYDIKIWS 789

Db 971 ENGLVAEPAQHAPI--RVSPLSKIALPSRAHLVLDIPKPLDRPTLAA-----FS 1019

QY 790 PLEESR 795

Db 1020 PVTKGR 1025

RESULT 9

T15348

hypothetical protein B0350.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999

C:Accession: T15348

R:Gattung, S.

A:Submitted to the EMBL Data Library, February 1996

A:Description: The sequence of C. elegans cosmid B0350.

A:Reference number: Z18332

A:Accession: T15348

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-5170 <GAT>

A:Cross-references: EMBL:U50071; NID:g1208871; PID:g1208877; PIDN:AAA93447.1; CESP:B0

A:Gene: CESP:B0350.1

A:Genetics:

A:Introns: 48/1; 5039/3; 5116/3

Query Match 4.5%; Score 202; DB 2; Length 5170;

Best Local Similarity 20.7%; Pred. No. 0.024;

Matches 125; Conservative 84; Mismatches 248; Indels 146; Gaps 24;

QY 191 RRAATSVACIPPIPYLLAVGCSDDSVRIYDRMLGTRATGNAGRGTTGMVARFIPSHLN 250

Db 520 RPISSEPAVSREPSHIATETTTTTRDYDQEPKADQSSARES-----VRDLVSEEH 572

QY 251 NKSCRVTSIYSEDGQEIIVSYSDIYLFDPKDDTA--RELKTPSAER-----REE 301

Db 573 HFACDSERLSEPAQSEPEPVETHAESQFSLVETTTTAVTREFYDEDEQEQASSRATKER 632

QY 302 LRQPPVKRLRLRGDWSDTGPRARP--ESERERDGEQSPVSLMORMDMLSRWFE--AS 357

Db 633 IEQSPVASER---SIVSTEHRQSPTQESQSLPTSEKNVHTVTETTTVTRCFEPIAS 689

QY 358 EVAQSNR--GRGSRPRGGTSQSDISTLTPVSPSP--DLEVSETAMEVDTPAEQFLQPTST 414

Db 690 ELBHAREFEQGNDRFRSVSPVDAEQVDVPHSPASHAESEQVPSPHPIHVETTTT 749

QY 415 STMSAQAHSTSSP--TESPHSTPLLSGPDSEQRQ-----SVEASGHTTHQSDNNN 463

Db 750 TTVTREFQDEYPRPESP--AEIFFIPSPSQSEQSEPHIVKETTITTTTRELDEPEKGN 807

QY 464 EKLSRPGT-----GEPVLS---LHSTGTTTSTIKLNTDEWSSIASRSGIGS 511

Db 808 VTFSPAPSSHAESERQVPSFVSHQYPPHVETTTTNTVTSNIYDDEDNVPSSEDPATQ 867

QY 512 HCKSEGESESFVQSS-----SVQPPPEGDS----- 535

Db 868 HFQ---QSEISVHRSHDPSVEESDGEGLGSKVLGFAKKAGMAGVGVAAAPVALAAGAKA 924

QY 536 -----ETKAPESSEEDYTKYQF--GVSAENPVENHINTQSDKKTAK 575

Db 925 AYDAFEKDEDEDETSHSPSPVPEYQSE---QYQDDSAQSSHTDFEHM----- 970



QY 576 PLDSNGERDNLDRSCGVPESASSEKAKEPE-TSDQTSTESATNNTNPE----- 628  
Db 971 PESPIHEKETEFEDHS--HPESPVLSEKERERHQVTSSTTTTIVTREVNDPEQEHQ 1028  
QY 629 -----POQTEATGSAHEETSTRSALQDQDSDSDPVLPGARYRAGDRRRSAVAR 682  
Db 1029 GPHSPAPSSHTAEAPHIVETTTTITVTRFEOPEE-----LEYKOEDNSRKSPSSH 1081  
QY 683 IOE 685  
Db 1082 SOE 1084

RESULT 10  
t29919  
hypotheical protein ZC449.5 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T29919  
R:Latreille, P.; Gattung, S.  
submitted to the EMBL Data Library, November 1995  
A:Description: The sequence of C. elegans cosmid ZC449.  
A:Reference number: Z20708  
A:Accession: T29919  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-550 <LAT>  
A:Cross-references: EMBL:U41510; PIDN:AAA82634.1; CESP:ZC449.5  
C:Genetics:  
A:Gene: CESP:ZC449.5  
A:Introns: 36/1; 73/3; 86/3; 110/3; 132/2; 469/3; 504/2

Query Match 4.4%; Score 199; DB 2; Length 550;  
Best Local Similarity 19.3%; Pred. No. 0.0019;  
Matches 102; Conservative 80; Mismatches 219; Indels 128; Gaps 16;

QY 249 LNKSCRTSLCYSEDGQELVSYSSDYIYLFDPKDTARELK--TPSAERREELQPP 306  
Db 116 LENKCLLTR-----ARRHAKSTTPATDDDDSEETTT 148  
QY 307 VKRLRLGDSDTGPRAPSERERDEQSPNVSLMQMSDMLSRWFEEA-SEVAQSNRG 365  
Db 149 VKSHKKSTTDAKPTTEEPASTKKKTESKKA VKTTTAAAPESAETPTSEEN 208  
QY 366 RGRSRPRGTSQSDISLTPVPSPDLVESETAMEVDTPAQFLQPTSTSTMSAQAHST 425  
Db 209 KTTSEESGNVSAEVTTSSTEVSEEQIVKSTEAETEEK-----STEAEEAATTS 262  
QY 426 SPTSEPHSTPLSPDSEQSQVSEAGHHTHQSDNNNEKLSPKPGTGPVLSLHYSTEG 485  
Db 263 SETSE-----TTERSEE-----EETTKPA-----KTHKSKT 290  
QY 486 TTTSTIKLFTDEWSSIASRRGIGSHCKSEGOEESFVQSSVQPPGDSSETKAPESSE 545  
Db 291 TTTT-----EASTKDSK---SHKTKEEKVTTTPEPTEESSESPTEEPQSSE 340  
QY 546 DVTKYQGV--SAENPVENHINTQDKFTAKPLDLSNGERDNLDRSCGVPESASSE 603  
Db 341 STPTSEESTEQSTEK-----KKEDKKDKDKKOKKED-----SEEDDDDK 384  
QY 604 KAKEPETSQDSTESATNNTNPEQFQTEATGSAHEETSTRSALQDQDSDSDPVL 663  
Db 385 KSKSSSSDSKDEKSTSSDSSEASSEKTVETGK-----KTLFSGPPDSDEDDDD--- 436  
QY 664 IPGARYRAGDRRSARAVARQIOEFFRRRKERK---EMEELDTNIRPLKMYKGRNRS 720  
Db 437 -----EGAGADEAFFSEKQATEAPAELSRTTVAVKSKMKKSGENFTP 482  
QY 721 TWIKENFANFVMSGDCGHIFIDWRHTABHMLLEADNHNVCLOP 769  
Db 483 LVI-----AGVFVSLGAGVYMFNKKERENLSTVDERELTIAP 524

## RESULT 11

S38170  
SRP40 protein - yeast (Saccharomyces cerevisiae)  
N:Alternate names: protein YKR092c; protein YKR412a  
C:Species: Saccharomyces cerevisiae  
C:Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 21-Jul-2000  
C:Accession: S38170; S40645; S37702  
R:Baladron, V.; Ballesta, J.P.G.; Bou, G.; del Rey, F.; Esteban, P.F.; Garcia-Cantale  
submitted to the Protein Sequence Database, March 1994  
A:Reference number: S38158  
A:Accession: S38170  
A:Molecule type: DNA  
A:Residues: 1-406 <BAL>  
A:Cross-references: EMBL:Z28317; NID:g486580; PID:g486581; MIPS:YKR092c  
A:Experimental source: strain S288C  
R:Bou, G.; Esteban, P.F.; Baladron, V.; Gonzalez, G.A.; Cantalejo, J.G.; Remacha, M.;  
Yeast 9, 1349-1354, 1993  
A:Title: The complete sequence of a 15 820 bp segment of Saccharomyces cerevisiae chr  
A:Reference number: S40644; MUID:94205265; PMID:8154186  
A:Accession: S40645  
A:Molecule type: DNA  
A:Residues: 1-406 <BOU>  
A:Cross-references: EMBL:X73541; NID:g450550; PIDN:CAA51946.1; PID:g450552  
A:Experimental source: strain S288C  
R:Lalo, D.; Carles, C.; Sentenac, A.; Thuriaux, P.  
submitted to the EMBL Data Library, May 1993  
A:Description: Interactions between three common subunits of yeast RNA polymerases I  
A:Reference number: S37702  
A:Accession: S37702  
A:Molecule type: DNA  
A:Residues: 1-399, 'N', 401-406 <LAL>  
A:Cross-references: EMBL:L11275; NID:g295670; PID:g295671  
C:Genetics:  
A:Gene: SGD:SRP40  
A:Cross-references: SGD:S0001800; MIPS:YKR092c  
A:Map position: 11R

Query Match 4.4%; Score 198.5; DB 2; Length 406;  
Best Local Similarity 23.2%; Pred. No. 0.0014;  
Matches 89; Conservative 56; Mismatches 172; Indels 67; Gaps 9;

QY 386 VPSSPDLEVTAMEVDTPAEQFLQPTSTSTMSAQAHSTSTSPHSTPLLS----- 438  
Db 8 VDEVPKLSVKEKEEEKSS 67  
QY 439 --SPDSQRQSVSEAGHHTHQSDNNNEKLSPKPGTGPVLSLHYSTEGTTTS----- 489  
Db 68 SDSSESSSESEDETK 127  
QY 490 -----TIKLNFTDEWSSIA--SSRGIGSHCKSEGOEESFVQSSVQPPGDS 535  
Db 128 KRARESDNEADAKETKKAKTEPESSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 187  
QY 536 ETKAPE-----SSEDVTKYQGVSAENPVENHINTQDKFTAKPLDLSNGERDNLN 588  
Db 188 ESDESDSQSSSDSD-- 245  
QY 589 LDRSCGVPESASSEKAKEPETSDQTSATNNTNPEQFQTEATGSAHEET--ST 646  
Db 246 ---SSGSDSSSDS 302  
QY 647 RDSALQDQDSDPVLPGARYRAGDRRSARAVARQIOEFFRRRKERKEMEELDTNIR 706  
Db 303 NESTPSASSSSANKLNIPAGTDEIKEGORK-----HFSRVDRSKINFEAWE----- 349  
QY 707 PLVKMYKGRNRSRTMIKEANFWG 730  
Db 350 -LTDNTYKG-----AAGTWG 363

RESULT 12  
T34513  
hypotheical protein ZK783.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
R:Accession: T34513  
R:Favella, A.; Vaudin, M.  
Submitted to the EMBL Data Library, August 1994  
A:Description: The sequence of C. elegans cosmid ZK783.  
A:Reference number: Z21536  
A:Accession: T34513  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-3507 <FAV>  
A:Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021; CESP:ZK783.1  
A:Experimental source: Strain Bristol N2; clone ZK783  
C:Genetics:  
A:Gene: CESP:ZK783.1  
A:Map position: 3  
A:Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/2; 3504/1

Query Match 4.4% Score 198; DB 2; Length 3507;  
Best Local Similarity 20.6%; Pred. No. 0.024;  
Matches 160; Conservative 112; Mismatches 282; Indels 222; Gaps 33;  
QY 77 ISNPYRKVLTTIRSGHRANIFSAKFLCTNDKQIVSCGDGVIFYTNVEODATNRQOC 136  
DB 1994 ISSTSTKMTSSKSPENVTMSSSEPVSTSSKSTTASSTV---SSTPSSSSSEAP 2049  
QY 137 FTCHYGTYEIMVNDPYFLSCGEGDGTVMFDTRIKTSCTKEDCKDDILLINCRRAATS 196  
DB 2050 LTSSPATTTT---SSVKSSTTPKESSESEITVKKLSKSP 2089  
QY 197 VAICPPPIYLVAGCSDSVRIYDRMLGTRATNYAGRTGMVARFIPSHLANKSCR 256  
DB 2090 V-----TESVK-----SSPTSTTSQSSTVTPETSKSTVLSSEAPV 2128  
QY 257 TSL-----CYSEGGQELVYSDDIYVLPDKDARELTPAEERRE----- 300  
DB 2129 TSTPTTEVHTSETKPSLASSTGDTNSTTPTSSLASVKSSTAPEGTSASVAPVKLS 2188  
QY 301 ---ELRQPPVKRL-----RLRGDWSDFGRARPDSERE-----RDGQSPNVSLMQ 343  
DB 2189 LSPDVQSSTTFDQATSTVQASSTSGTSVKSSTSEPHVTKLSITSSNPSSVPVTS 2248  
QY 344 RMSDMLSRWFEEASEVAQSNRGRS-RPRGTSQ-----SDISTLPTVP 387  
DB 2249 PKS---TPTVPESTEQPTSTTPSGQSLAPMNSSEVLTTSEPHVLSLSPDVQSSTTP 2305  
QY 388 S-----SPDLEYSETAMEVD-----TPAQFLOP-----STSSMSAQAHS--- 423  
DB 2306 NNLSESTVETPKTSSEVLSNSEEPTTEAPTLLSPDILSTTNLSQSTYSTEDRSI 2365  
QY 424 TSSPTSPHSTPLL-----SSPD---SPQSVASGHHHTHQSDNNNEKL---SP 468  
DB 2366 SSENSEKPTSAPELVTSSVTHVASSSPDPTSESPDPLTGSSTENIPASSKQITSTP 2425  
QY 469 KPGT---GEPVLSHYSTEGTTSTI---KLNFTEWSS---IASSRGIGSHCKSEGE 519  
DB 2426 TPDUTTATSEPTKTSMSPDLTSTTNVLSSESTTPSSKSPVSSSTEGISVVTST---E 2482  
QY 520 ESFVPOSSVQ-----PPEGDSETKAPEESEDVTK-----YQGVSAEN- 558  
DB 2483 FKVPSTISSVLEEDLTFTTSPILEETTTASETSEPTLEDLSVSVRIHELTTSEN 2542  
QY 559 PVENHINITQSKFTAKPLDSNGERDNLDRSCGVPESASSEKAKEPE----- 609  
DB 2543 PRESESTTTSSSE---SSKPSQEPAG-----ILTSTVVVPTSSVSLITASEIAITSTNTPK 2595  
QY 610 -----TSDQSTESATNENN-TNPEQFQTEATGPSA-----HEETSTRDSAL--- 651  
DB 2596 QGRTPTTSPKSLVKSTTSPSTVTSSEPSSTKRTVSTVSTTTTTEETTTSESILTA 2655  
QY 652 -----QDTRDSDDDVPLIPGARYAGPDRRSASAVARIOEFFRRRKERKEMEELDT 701

Db 2656 APSKPTSTESSEAPT-TPAKTSETKPSNVST-----SRKSTENVET 2698  
RESULT 13  
T06310  
hypothetical protein Fl1C18.80 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 22-Oct-1999  
R:Accession: T06310  
R:Revan, M.; Terryn, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.;  
ewes, H.W.; Mayer, K.F.X.; Schueller, C.  
Submitted to the Protein Sequence Database, April 1999  
A:Reference number: Z15589  
A:Accession: T06310  
A:Molecule type: DNA  
A:Residues: 1-852 <BEV>  
A:Cross-references: EMBL:AL049607; GSPDB:GN00062; ATSP:Fl1C18.80  
A:Experimental source: cultivar Columbia; BAC clone Fl1C18  
C:Genetics:  
A:Gene: ATSP:Fl1C18.80  
A:Map position: 4  
A:Introns: 34/3; 102/3; 163/2; 205/3; 269/3; 282/3; 564/3; 602/3; 622/2; 639/3; 667/3

Query Match 4.3% Score 193; DB 2; Length 852;  
Best Local Similarity 21.9%; Pred. No. 0.0071;  
Matches 129; Conservative 76; Mismatches 224; Indels 160; Gaps 26;  
QY 162 EDGVRWFDTRIKTSCTKEDCKDDILLINCRRAATSVACPPPIYLVAGCSDSVRIYDR 221  
DB 365 ESSSVKQADLS-KSDDIKEETEPAEILLDSKDLTS---PPV-----DSSV----- 405  
QY 222 RMLGTRATNYAGRTGMVARFIPSHLANKSCRVTSLCYSEGGQELVYSDDIYVLP 281  
DB 406 ---TAATSS-----ENKNSV-----QILPKSTSG----- 428  
QY 282 PKDGTARELTKTPSAERREELRQPPVKRLRLKGDWSDTGPRAPESEERDEGEQSPNVSL 341  
DB 429 ---DETA-NVSSPSMAEELPEQSV-----KKTANOKKESSTE---EVKPSASI 471  
QY 342 MORMSDMLSRWFEEASEVAQSNRGRSRPRGTSQSDISTLPTVP-----SSPDLEV--- 394  
DB 472 AT-----EEVSEPTSEPO-VTKKSKKVASSSKTKPTVPPSKKSTSETKVAKQ 520  
QY 395 ---SETAMEVDTPAEQFLQSTSTMSAQ-HSTSSPTESP-HSTPLLSPDSQ 444  
DB 521 SKKVVGSNDQAEKTPKEKKKPGRAIDESLHTSGDNEKPAVSSGKLASKKKEA 580  
QY 445 ROSVEASGHHHTHQSDNNNEKLSPKPGTGEVPLSHYSTEGTTSTIKLNFTEWSSIAS 504  
DB 581 KQTV-----ESPNSNTKRRSLGQGA-----SGESLVGSRKIV---WPMDOA 622  
QY 505 SRRGIGSHCKS-----EGQESFVPSQSVQPEGDSETKAPES-----EDYT 548  
DB 623 YKGVVSTDAKKHLYIYDGDQELLYLKQKWSPLDESELSDQDEEAADQGTQEDAS 682  
QY 549 KYQEGVSAAE---NPVENHINITQSKFTAKPLDSNGERDNLDRSCGVPESASSEK 604  
DB 683 TVGSGAGSSKAKATPASKSKTSQDDKTASKSKDSKEASR-----EEEEASSEE 730  
QY 605 AKPEPESDOT---STESATNENNTNPEQFQTEATGPSAHEETSTRDSALQDTRDSD 661  
DB 731 ESEEEPPRTVKGSGSRSKKDISSVSKSGKSKASSKKKEEPSKATSS-----KSKSGP 785  
QY 662 VLIPIGARYAGPDRRSASAVARIOEFFRRRKERKEMEELDTLNIRRLPVK 710  
DB 786 VKSVPAKSKTGKAKSGSASTPA-SKAKESASESESEETPKPEPATK 833

## RESULT 14

S48478  
glucan 1,4-alpha-glucosidase (EC 3.2.1.13) - yeast (Saccharomyces cerevisiae)  
N:Alternate names: extracellular glucanase; mucin-like protein MUC1; protein YIR01  
C:Species: Saccharomyces cerevisiae

C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 12-Nov-1999  
 C:Accession: S48478; A26877; B26877; S27281; JC6123  
 R:Rowley, K.  
 Submitted to the EMBL Data Library, October 1994  
 A:Reference number: S48478  
 A:Accession: S48478  
 A:Molecule type: DNA  
 A:Residues: 1-1367 <ROW>  
 A:Cross-references: GB:Z47047; EMBL:Z38061; NID:G603997; PID:G763364; GSPDB:GN00009; MIM  
 R:Yamashita, I.; Nakamura, M.; Fukui, S.  
 J. Bacteriol. 169, 2142-2149, 1987  
 A:Title: Gene fusion is a possible mechanism underlying the evolution of STAL.  
 A:Reference number: A91831; MUID:87194600; PMID:3106330  
 A:Accession: A26877  
 A:Molecule type: DNA  
 A:Residues: 1-242 <YAM>  
 A:Cross-references: EMBL:M16164; NID:G172522; PIDN:AAA35014.1; PID:G172525  
 A:Accession: B26877  
 A:Molecule type: DNA  
 A:Residues: 762-1331 <YA2>  
 A:Cross-references: EMBL:M16165; NID:G172523; PIDN:AAA35015.1; PID:G172526  
 R:Pardo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.  
 FEBS Lett. 239, 179-184, 1998  
 A:Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Sacchar  
 A:Reference number: S27281; MUID:89031230; PMID:3141213  
 A:Accession: S27281  
 A:Molecule type: DNA  
 A:Residues: 1-31 <PAR>  
 A:Cross-references: EMBL:X13857; NID:G4551; PIDN:CAA32069.1; PID:G4552  
 R:Lambrechts, M.G.; Bauer, F.F.; Marmur, J.; Pretorius, I.S.  
 Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996  
 A:Title: Muc1, a mucin-like protein that is regulated by Msi10, is critical for pseudohy  
 A:Reference number: JC6123; MUID:96322323; PMID:8710886  
 A:Accession: JC6123  
 A>Status: nucleic acid sequence not shown  
 A:Molecule type: DNA  
 A:Residues: 1-1367 <LAM>  
 A:Cross-references: GB:U30626; NID:G1304386; PIDN:AAC49609.1; PID:G1304387  
 C:Genetics:  
 A:Gene: SGD:MUC1; STA2; MAL5; DEX2; SGD:S0001458  
 A:Cross-references: MIPS:YIK019c; SGD:S0001458  
 A:Map position: 9R  
 C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase  
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein  
 F:5-21/Domain: transmembrane #status predicted <TM>  
 F:1350-1366/Domain: transmembrane #status predicted <TM2>

Query Match 4.2%; Score 190.5; DB 1; Length 1367;  
 Best Local Similarity 22.3%; Pred. No. 0.018;  
 Matches 124; Conservative 64; Mismatches 270; Indels 99; Gaps 18;

DB 151 PNDPTFLSCGDCGTVRWFDTRIKTSCTKEDCKDDILNCRRATSAICPPYPLAVG 210  
 DB 272 PTPPTTSCVKEKPPPHDFT---TPCTKK--KTTSTKCTKKT---PVPTPSSST 321  
 QY 211 CSPDSVRIYDRRLMGLTRATNYAGRTGMVARFIP---SHLNKSCRVTSLCYSEDQ 266  
 DB 322 TESSAPVPPTSSSTTESSAPVPTSSSTTESSAPVPTSSSTTESSAPVPTSSSTTESSA 381  
 QY 267 EILVSYSSDIYFLDPKDDTARELKTSAERRELRQPPVKRLRLRGDWDSDTGPRAPE 326  
 DB 382 PVTSSTT-----ESSAPVPTSS-----STTESSAPV 410  
 QY 327 SERERCEQSPNYSMLQRMDSLSRWFEEAEVAQSNRGRSRP-RGGTQSQSDISTLPT 385  
 DB 411 TSSTTESSAP-----VTSSTTESSAPVPTSSSTTESSAPVPTSSSTTESSAPVPT 460  
 QY 386 VPSSPLEVSE---TAMEVDTPAEQFLQPTSTSTMSAQHSTGPTES---PHSTPLLS 439  
 DB 461 -PSSSTTESSAPVPTSSSTTESSAPVPTSSSTTESSAPVPTSSSTTESSAPVPTSSST 519  
 QY 440 PDSEQRQSVASGHHHQHQS-----DNNNEKLS---PKPGTGPVLSLHYSTGTTSTIK 492

DB 520 TESSAPAPTSSSTTESSAPVPTSSSTTESSAPVPTSSSTTESSSTPVTSTTESSA 579  
 QY 493 LNFDEWSSIASSSRGIGSHCKSEGESEF---VPOSSVQPPGEGDSEKAPESSEDVTK 549  
 DB 580 PVPTPSSSTTESSAPVPTPSSSTTESSAPAPTSSSTTESSAPVPTSSSTTESSAPVP 639  
 QY 550 YQEGVSAEN---PVENHINITQSKFTAKPLDSNGERNDLNDRCGVPPEASASSEKAK 606  
 DB 640 TPSSSTTESSAPVPTPSSSTTESSAPVPTPSSS-----TTESSAPVPTSSSTESS 692  
 QY 607 EPETSDQFSTESA-----TNENNTNPEQFQTEAT-----GPSAHEETSTRDSA 650  
 DB 693 APVTSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPV 752  
 QY 651 LQDITDDSDDDPVLPGA 667  
 DB 753 TSSTTESSAPVPTPSS 769

RESULT 15  
 A41859  
 Iga-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus influ  
 C:Species: Haemophilus influenzae  
 A:Variety: strain HK715  
 C>Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 08-Dec-2000  
 C:Accession: A41859  
 R:Poulsen, K.; Reinholdt, J.; Killian, M.  
 J. Bacteriol. 174, 2913-2921, 1992  
 A:Title: A comparative genetic study of serologically distinct Haemophilus influenzae  
 A:Reference number: A41859; MUID:92234949; PMID:1373717  
 A:Accession: A41859  
 A>Status: preliminary; not compared with conceptual translation  
 A:Molecule type: nucleic acid  
 A:Residues: 1-1702 <POD>  
 A:Cross-references: GB:M87489; NID:G148906; PIDN:AAA24966.1; PID:G148907  
 A:Experimental source: strain HK715  
 A:Note: Sequence extracted from NCBI backbone (NCBIP:97282)  
 C:Superfamily: Iga-specific metalloendopeptidase  
 C:Keywords: hydrolase; metalloproteinase

Query Match 4.2%; Score 188; DB 2; Length 1702;  
 Best Local Similarity 18.9%; Pred. No. 0.032;  
 Matches 150; Conservative 120; Mismatches 311; Indels 212; Gaps 33;

QY 67 LSGSDDTKLVISNPYSRKVLTIIRSGHRANFISAKFLPCTNDKQIVSCGDGVITYTWE 126  
 DB 773 ITASDNKAVHIG--YKAGDTVCYRSDYTG-----YVCTTDK--LSDKALNSFNATNVS 822  
 QY 127 QDAETNRQCOFTCH---YGT---TYEIMTVPNDPYTLSCGEGDGTVRWFDTRIKTSC 177  
 DB 823 GNYNLSGNANFVLGKANLFGTISGTGNSQVRLTENSHTWLTG---DSNVNQLNLD----- 874  
 QY 178 TKEDCKDDILNCRRATSAICPPIP-----YLLAVGSDSVRIYDRRLMGLTR 227  
 DB 875 -----KGTHLNAQNDANKVTTNTLVNLSLGSNGSPYLL---TDLNKGQDKVVVTKS 925  
 QY 228 ATGNY-----AGRTTGMVARFIPSHLNKSCRVTSLCYSED---GOETLVSYSSDYI 277  
 DB 926 ATGNFTLQVADKTGETPKNELTLFDASNATNENNLNLSLVGNTVDLGAWKYKLRNVNGRY- 984  
 QY 278 YLFDPK-----DDTAR-----ELKTPSAERRELRQ-----PPVKRLRLRGDWS 318  
 DB 985 DLYNPVEKRNQVDTNTITPNNIQADVPSVPSNNIEARVETVPVPPAPATPSETTET 1044  
 QY 319 TGPRAPES-----ERERDG-----EQSPNYSMLQRMDSLSRWFEE----- 354  
 DB 1045 VAENSKQESKTEVKNQDQDATTETAQNGEVAEAKPSVKANTQTNEVAQSGSETETQTE 1104  
 QY 355 --EASEVAQSNRGRSRPRGGTQSQSDISTLPTVPS-----SPDLEVSETA 398  
 DB 1105 IKETAKVEERAKVEEKAKVEKDEIQEAPQWASETSPKQAKPAPEVSTDKVETQ 1164  
 QY 399 MEVDTPAEQFLQPTSTSTMSAQHSTSSPESHPTLLSSPDSEQRQSVASGHHHTHQ 458

358	Db	ElVrIfSVNkEEVSReIvStSt-----TAPSPRIVeGtKtKtQVlKQePvEtGvEHkD	408
334	QY	EQS-----PNVSlMORMSdMLSRWFEEASVAQSNRGRGRPRGGTSQSDISlTPTV--	386
410	Db	VQSGAlVErPAIQ-----PElPEAVVSDKGePEVQ-----TLPEAVV	446
387	QY	-----PSSPDlEVSEtA-----MEYDTPAEQfLQlPSTStSMsAQ	420
447	Db	TdKGETEvQeSPdTVVSDKGePEQVAPlPEYKGNlEQVKEtPVeKtKQeGPeKtEEV-	505
421	QY	AHSTSSPTeShTlLLSSPDSeQoRQSVeASGHtH--HQSDNNNEKlSPKPGtGEPVlS	478
506	Db	-----PVKPTETPVNPNegTegTSQeAENPVQAPAEStNSEKVS-----	549
479	QY	LHyStEGtTtStIKlNfTDewSsIASsRGlSGHCKSeQeSfVQoSVoPPeGdStK	538
550	Db	---dTSKktGEVSNpSDStTSvGESnkPEHdSKNENSEKt-VEEVPVNPNEGtVegT	605
539	QY	APeESeSdYtKQeGVSAENPVENHlNlTQSDKfTAKPLdNSNGERNdLDRSCGVPEE	598
606	Db	SNQETeKPVQAPAE-----TQTN--SGKlANETGVSNKpSDSKPPV--EE	648
599	QY	SASSEK---AKePETSdOTStESATNENNTNPfQTeATQPSAHEETStRdSLQDtd	655
649	Db	SNQPeKNGtATPeNSGNTtSEN---GQTePpSNGNStEDVSStESNTSNGNEEKQ	704
656	QY	DSDDDP	661
705	Db	ENELDP	710

**RESULT 17**

H64106  
Iga-specific metalloendopeptidase (EC 3.4.24.13) type 1 - Haemophilus influenzae (str N; Alternate names: immunoglobulin A1 proteinase type 1  
C:Species: Haemophilus influenzae  
A:Variety: strain Rd KW20  
C:date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 08-Dec-2000  
C:accession: H64106; A41500  
R:Rietschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage  
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman  
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagan, N.S.M.  
Science 269, 496-512, 1995  
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter  
A; title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A:Reference number: A64000; MUID:95350630; PMID:7542800  
A:Accession: H64106  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA

A: Experimental source: strain Rd KW20  
 R: Grundy, F.J.; Plaut, A.G.; Wright, A.  
 Infect. Immun. 58, 320-331, 1990  
 A: Title: Localization of the cleavage site specificity determinant of Haemophilus influenzae type 1  
 A: Reference number: A41500; MUID:90129281; PMID:2105270  
 A: Accession: A41500  
 A: Molecule type: DNA  
 A: Residues: 1-377 <GRU>  
 A: Cross-references: GB:X59800  
 A: Experimental source: strain Rd KW20  
 A: Note: the authors translated the codon TGG for residue 319 as Thr  
 C: Function:  
 A: Description: this protecnase is classified as type 1 because it cleaves at a proline  
 C: Superfamily: IGA-specific metalloendopeptidase  
 C: Keywords: hydrolase; metalloproteinase

Best Local Similarity	19.48; Pred. No. 0.041;	
Matches	153: Conservative	113: Mismatches
	310:	Indels
	214: Gaps	33:

QY 67 LSGDDTKLVISNPYSRKVLTTIRSHRANIFSAKFLPCTNDKQIVSCSDGVIFYTNVE 126

Db 773 ITASNAKVHIG--YKAGDTVCVRSDYTG-----YVCTTDDK--LSDKALMSFNATNVS 822  
QY 127 QDAETNRQCFCH-----YGR-----TYEIMTVPNDRPTFLSCGDBGVVRWFDTRIKTSC 177  
Db 823 GNVNLGNANFVLKANLFGTISGTGNSQVRLTENSWHHLTG-----DSNVQNLND----- 874  
QY 178 TKEDCKDILLNCRRAATSAVLCPIP-----YLLAVGSDSSVRIYDRMLGTR 227  
Db 875 -----KGHIHLNAONDANKVVTYNTLTVNSLSGNGSFYLL-----TDLNKGQDKVVVTKS 925  
QY 228 ATGNY-----AGRGTTGMVAREIPSHLNKSKRVTSCLCYSED-----COETLVSYSSDI 277  
Db 926 ATGNFTLQVADKTGPTKNELTFLDASNATNENLNVSLGVNTDLGAWKYKLRNVNGRY- 984  
QY 278 YLDPDK-----DDTAR-----ELKTPSAERREBLRQ-----PPVKRLRLRGWSD 318  
Db 985 DLYNPEVKNQVTTNITPNNIOADVPSPVSNNEIARVETPPVPPAPATSETTET 1044  
QY 319 TGPRAPESERERDEQSPNVSLMORMSDMLSRWFEEA-----GEVAEAKPSVKANTQTNEVAOQSGSETE 1099  
Db 1045 VAENSKQESKTVEKNEQDATETTAQN-----GEVAEAKPSVKANTQTNEVAOQSGSETE 1099  
QY 363 -----NRGRSRPRGGTQSIDSITLPTVPS-----SPDLEVSETAMEY 401  
Db 1100 TQTEIKETAKVKEEAKAKYKDEIOAPQWASSETSPKQAKPAPKEVSTDTKVEETQVQA 1159  
QY 402 DTPAEQFLOPSTSTMSQAQAHSTSPPTSPHSTPLSSPSEORQSVASGHHTHQSDN 461  
Db 1160 -----QFOTOSTVVAEAT-----SPNSKPAEETOPSEKTNAPVPPVVSKNQTN 1206  
QY 462 NNEKSPKPGTGPVLSHYSTEGTTSTIKLNFOTDEWSSIASRRGIGSHCKSEGOEES 521  
Db 1207 T-----TDQP-----TEREKTAKEVTEKTOEPQVQSOA-----SPKQESQET 1244  
QY 522 FVQSSVOPPEGSETKAPESSEDVTKQGVSAENPV-ENHIN-----ITQSDKFTA 574  
Db 1245 VOQAVLESENPTVYNAEVOAQLOTQTSATVSTKOPAPENSINTGSATATETAEKSD 1304  
QY 575 KPLDNGSERNDLNDRSCGVPEEASSEKAKEPETSDQTSATNENNTNPEPQFOTE 634  
Db 1305 KP-----QTETRASTEDASQHKANTYADNSVANNSSOPKSRRRS 1346  
QY 635 ATGPSAHEETSTDSALQTDSDDDPVLPGARVAGPGRDRSARVARIQEFRRKERRK 694  
Db 1347 ISQP-----QETSAEETAASTDET-----TIADNSKRSP-KNRS-----RRSVRS 1387  
QY 695 EMEELDTLIRPLVMYKGRNKRMTIKNEANFVMSGDCGHIFI---WDRHTA 751  
Db 1388 E-----PTVNGSDRSTVALRDLTSTNTNAVISDAMAKAQFVALNVGKAVS 1433  
QY 752 EHLMLLEADN 761  
Db 1434 OHISOLEMNN 1443

RESULT 18  
T29612  
hypothetical protein D1014.5 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T29612  
R:Du, Z.; Leimbac, D.  
submitted to the EMBL Data Library, March 1996  
A:Description: The sequence of C. elegans cosmid D1014.  
A:Reference number: Z20652  
A:Accession: T29612  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-727 <DUZ>  
A:Cross-references: EMBL:U53180; PIDN:AAA96284.1; GSPDB:GN00023; CESP:D1014.5  
A:Experimental source: strain Bristol N2; clone D1014  
A:Genetics:  
A:Gene: CESP:D1014.5

A:Map position: 5  
A:Introns: 52/1; 379/2; 442/2; 579/3; 618/2; 647/3; 676/2; 707/3  
Query Match 4.0%; Score 183; DB 2; Length 727;  
Best Local Similarity 23.5%; Pred. No. 0.02; Indels 88; Gaps 14;  
Matches 89; Conservative 47; Mismatches 155; Indels 132; Gaps 28;  
QY 332 DGEQSPNVSLMORMSDMLSRWFEEASEVAQSNRGRSRPRG-----GTSQSDISTLPTVP 387  
Db 43 DSKASPVSDRQOYSD-----QOKEVNSNSNTESKNAGSNPGSSSETADQTAPEP 95  
QY 388 SSPDLEVSETAMEYDTPAEQFLOPSTSTMSQAQAHSTSPPTSPHSTPLSSPSEORQSV 447  
Db 96 -KPKVSDDNAGSAETIESNGIEKIPS-----NDTSPFDSAVVP--SGPAQEEDEA 145  
QY 448 VE-----ASGHHTHQSDNNNEKLSPKPGTGPVLSLH-----YSTEGTT----- 487  
Db 146 VEEKKVEEHEENHNGSDEAEKESGDAVTTTSLMEDREGNDPDKDESGTIVTSNE 205  
QY 488 -TSTIKLNFDEWSSIASSS---RGISHCKSEGOEES-----FVP 524  
Db 206 ETTTTEKNKVADEAPSKAEISQASQSDSHVKEGESETTTTSDSATTEKKTNGAELVE 265  
QY 525 QSSVQPPGEGSETKAP-----ESSEDVTKQGVSAENPVENHINTQSDKFTAKPLDNS 581  
Db 266 KSETEKNGDSSTKSPVETSESSVDTTIEQLKSNDEE-----KQVDGNE 312  
QY 582 GERNDLNDRCGVPEEASSEKAKEPETSDQTSATNENNTNPEPQFQTEATGPSAH 641  
Db 313 KESTFRVLNEDHNEEIEGKTK-----KVTEETTTVETTERQNVDDTENTQTTTSSSY 368  
QY 642 -----EETSTRDSALQDIDD 656  
Db 369 NAKEEKSYRCKKVEWND 387  
RESULT 19  
F90073  
hypothetical protein SA2447 [imported] - Staphylococcus aureus (strain N315)  
C:Species: Staphylococcus aureus  
C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C:Accession: F90073  
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O-  
ma, A.; Mizutani-U, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.  
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
Lancet 357, 1225-1240, 2001  
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A:Reference number: A99758; MUID:21311952; PMID:11418146  
A:Accession: F90073  
A>Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-2271 <KUR>  
A:Cross-references: GB:BA000018; PID:gl3702612; PIDN:BA43752.1; GSPDB:GN00149  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: SA2447  
Query Match 4.0%; Score 183; DB 2; Length 2271;  
Best Local Similarity 21.1%; Pred. No. 0.087;  
Matches 147; Conservative 101; Mismatches 316; Indels 132; Gaps 28;  
QY 55 NTICWNTDGEYILSGSDDTKLVISNPYRKYVLTTR-----SGHRANIFSAKFL 103  
Db 939 NSIANSQSASTSKSDQSTSTSLSTSDSKMSSTSESLDSTSTSGSVGSLIAASQSVS 998  
QY 104 PCTND-----KQIVSCSGDGVIFVTNVEQDAETNRQCF-----TCHYGTTEIM-----T 149  
Db 999 TSTSDSMSTSEIVSDS-----ISTSGLSASDSKMSVSSMSVSSMSQSGTSESLSDSQST 1053  
QY 150 VPNDPPTF-LSCGEDGVTVWFDTRIKTSCTKCDKDDILNCRRAATSAVACPPIPYLA 208  
Db 1054 SDSDSKSLSTSQSGST---STSTSTASVSTSESQSTSGSMSASQSDSMS-----IS 1104

[illegible]

RESULT 20  
F89500  
protein T27A10.6 [imported] - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Nov-2001  
C:Accession: F89500  
R;anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology  
A:Reference number: A75000; MUID:99069613; PMID:9851916  
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/projects/C\_ele  
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
A:Accession: F89500  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-868 <STO>  
A:Cross-references: GB:chr\_X; PIDN:AAB52504.1; PID:g1326288; GSPDB:GN00028; CBSP:T27A10  
C:Genetics:  
A:Gene: T27A10.6  
A:Map position: X

	Query Match	4.0%;	Score 180.5;	DB 2;	Length 868;
	Best Local Similarity	20.8%;	Pred. No. 0.034;		
	Matches 142;	Conservative 91;	Mismatches 246;	Indels 203;	Gaps 35;
QY	129	AETNRCCQCTCHGVYTIETVNDPVTLSGCE----	DGTVRWFDRIKTSCTKEDCKD	184	
DB	78	AEIERE----DCHVSMIDPILDDNTFTT----	CGPREVLGITRLSDRIQTMC	124	
QY	185	DILLNCRRATSAVAICPPPIYYLAVGSDS-----	SVRIYDRRMGLGTRATGNYA	233	
DB	125	-----CRLSRDEFENCKDLIFNPIGLTRSTLIENDNQLINAIRIEER-----	NYV	170	
QY	234	GRGTTGWARFIPTSHLNKSCRVTSLCYSDGQEIILYSYSDSYILFDPKDDTARELTKTP	293		
DB	171	VR--FCDLAPRATASILSDEKVRPSTPAEATEDVPVSYQELD--TFSERODKQNLKVV	227		

294	QY	SAERREELRQPPVKRLRLRGDSWTDGQRPARESERERDEQSPNVSLQMRMSDMLSRWF	353
228	Db	KTQGAEE--APPTK-LRV-----PMHPFSRVEAAPVVEFQ--PAVPLIEIKSD-----	271
354	QY	EEASEVAQSNRGRGRSPPRGCTSQSDISILPVPSPDPLEVSETAMEVDTTAEQFLOPST	413
272	Db	--ADDI-----SEGALSPVAPQVTTTRLATVSLAUVI-----IPPTT	308
414	QY	SSTMSAQAHSTSSPTES-----PHSTPLLS-----SPOSEQRQSVEASGH	453
309	Db	STTEELPVQEQLPVTPEIPDEETTFPOETPVFSQOLVEDIVKKGISAQSEQK-AAQIFQQ	367
454	QY	HTHOSHDDNNKE-----LSPK--PGTCEPVLSL---HYSTEGTITSTIKLNTDEWSSIA	503
368	Db	SLNKLLQANAEKPIEAVSAPKAGPAPRQPNQLDDVNFENEOFPQPSGKHRFSKELLPLR	427
504	QY	SSSRGIGSHCKSEGESEFVQSSQVQPPEDSETKAP-----EESSESDVIKYQ	551
428	Db	KSADG-----EDF---SNLPVNTEGRRRSPITRHPVHDWDSDEDEDNDKIV	473
552	QY	EGVSAENPVENHINTQSDKETAPLDSNGSERNDLNDRCGVPSESASSEK-----	604
474	Db	SEINGS--VEMHFET-----LPPKKVQEN--EKNEKN-EKKNFIDESKDYDTONLMLKEQ	524
605	QY	-----AKEPETSOTSTE--SAYNENNTNPEQF-QTEATGSAHEETSTR-----	647
525	Db	IVKKIIKVVAHRFRPPTTTEETTEESTKTTPIRTRPTTTTTEETTRKIAN	584
648	QY	--DSALQDTPDDSDDDPVLI-----PCARYRAGPGBDRSARVAQIEFFRR--	690
585	Db	FFNSMQEYESPQRNOQAVQTQPRFRGRTTTPSAR-RDLPVTRVAETVNOQDNARYI	643
691	QY	-----KERKEMEDELDT	701
644	Db	SPENQVNEFTLEIKT	665

RESULT 21  
T46481  
hypothetical protein DKFZp434A025.1 - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
C:Accession: T46481  
R:Duisterhoef, A.; Lauber, J.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
Submitted to the Protein Sequence Database, January 2000  
A:Reference number: Z23035  
A:Accession: T46481  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1027 <AAR>  
A:Cross-references: EMBL:AL137755  
A:Experimental source: adult testis; clone DKFZp434A025  
C:Genetics:  
A:Note: DKFZp434A025.1

	Query Match	3.9%	Score 178;	DB 2;	Length 1027;
	Best Local Similarity	20.1%;	Pred. No. 0.058;		
	Matches 131;	Conservative	79;	Mismatches 258;	Indels 184; Gaps
QY	295	AEERREELRPPVKRLRLRGDWSDTGPR-----ARPESEGEREDGQSPNVSLMQMSDML	349		
		: :    :    :    :    :    :    :    :    :    :    :    :    :    :			
DB	246	AERLQRQDQEQAYLSLQHDHRRHPHQSHQSPPPQQRKSPKSFHAPKHAHEPAD--	303		
		: :    :    :    :    :    :    :    :    :    :    :    :    :    :			
QY	350	SRWFEEASVAQSNRGRSRPRGGTGSQSDISTLPTVPS-----	388		
		: :    :    :    :    :    :    :    :    :    :    :    :    :    :			
DB	304	-----RAREVEDRPKTNHSSPAQSKQGRVLEPPVPVPSSEFSGNGNSESVHPALQRP	358		
		: :    :    :    :    :    :    :    :    :    :    :    :    :    :			
QY	389	SPDLEVSETAM-----EVDTPAEQFLQ-----PSTSTMSAAQHSSTSS	426		
		: :    :    :    :    :    :    :    :    :    :    :    :    :    :			
DB	359	EPQVQWHLAKKNVSPVRSRHSFSDPKFAHHHLRQDPCPPSRSEVLQSQSDSKSE	418		
		: :    :    :    :    :    :    :    :    :    :    :    :    :    :			
QY	427	-----PTESPSTPLSSPDSEQRQSVFASGHHHTHQSNNN--	463		

Db 419 APDPTOKANRSRSDSEVPFPRVTRTSRSPVLRRDS-----PLOGSGQNSQAGQRNSTS 474  
 QY 464 -----EKLSPKPGTGEPVLSLHYSTEGTITSTIKLNTDEWSSIASSSRGIG-- 510  
 Db 475 SIEPRLWVERVKLVPRPGSG-----SSGSS-----NSGSPGSHFGSQSGGER 520  
 QY 511 ----SHCKSEGOEESFVPOSSVQPPGCDSETKAP--EESSEDVTKYQOEGVSAENPVENHI 564  
 Db 521 FVRSSSKSGSPSQL--ENAVKPEDKKEVFRPLKPAGEVDLTALAKELRAVEDVRPPH 579  
 QY 565 NITQSKDKFTAKPLDSSNGSRNDLNL-----DRSCGVPER--SASSEKAKEPETSDDTSTE 617  
 Db 580 KVTD---YSSSEESTIDEEDDVEQCADSTSPEDTRASSLNLNGETE---SVK 633  
 QY 618 SATNENNTPEQFQTEAGPNSAHEFTSTRDSALQDSTDSD-----DPVLI----- 664  
 Db 634 TMIVHDDVESEFAMTPSKEGTLIVRQTQSASTLQKHSSSFTPTIDPRLQLQIPSSGT 693  
 QY 665 -----PGARYRAGPD--RRSAVARIQEFFRRRERKEMEELDTLNRRLPKVMV 712  
 Db 694 TVTSVVGSCDWRPEAIRQDTRKGVSVNVNPTNTR-----PQSDTPEIRK----- 740  
 QY 713 YKGRNSRTMIKEANFWGAFYMGSDCGHIFIDWRHTAEHLMLLE--ADNHVNCQLQPH 770  
 Db 741 YKKRENSEILC--AALGWVNLIV-GTESG-----LMLLDRSCQGVYPLINRR 785  
 QY 771 PPDPILAGSGIDYDIKINSPLEESRIFNRKLADEVITNELMELETRNTIV 822  
 Db 786 RFOQMDVLEGLNVLTISGKKDKLVYLSWLNRKLHNDPEVEKKQGTIV 837

## RESULT 22

T48933  
 WD repeat domain protein - Arabidopsis thaliana  
 N:Alternate names: protein F14L2.80  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000  
 C:Accession: T48933  
 R:Jordan, N.; Bangert, S.; Wiedelmann, R.; Voss, H.; Unseld, M.; Mewes, H.W.; Rudd, S.;  
 submitted to the Protein Sequence Database, April 2000  
 A:Reference number: 225008  
 A:Accession: T48933  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1051 <OR>  
 A:Cross-references: EMBL:AL353818; GSPDB:GN00061; ATSP:F14L2.80  
 A:Experimental source: cultivar Columbia; BAC clone F14L2  
 C:Genetics:  
 A:Gene: ATSP:F14L2.80  
 A:Map position: 3  
 A:Introns: 36/3; 121/3; 205/3; 336/2; 668/3; 739/1; 773/1; 853/2; 894/2; 932/3

Query Match 3.9%; Score 178; DB 2; Length 1051;  
 Best Local Similarity 18.8%; Pred. No. 0.059;  
 Matches 128; Conservative 76; Mismatches 212; Indels 264; Gaps 26;

QY 42 KLEATLVHVDGCVNTICWNTDGEYILSGDDTKLVI--SNPYR-----K 84  
 Db 50 RLLATLRDFGVCNVRWAKNSRYVAGSDQVQIHERPFGSGTTEFGSGEAPDVENWK 109  
 QY 85 VLTIRSGHRANFSAKFLPCTNDKQIVSGSGGVPIYTNV----- 125  
 Db 110 AVMTLR-GHTADVVDLNMWSP--DDSLASGLDNTVHWNRTGMCITVLRLHLSLVKG 166  
 QY 126 -----EODAE----- 125  
 Db 167 TWDPIGSFASQSDDKTVIWRSTDGMMAHRTDGHNAKSLGSLGTFRRGLWSGPCHELTIT 226  
 QY 146 EIMTVPN-----PYTFL-----SCGEDGVVRWFDR 172  
 Db 227 HGQPKHSAVPLRGESVAYDFLGHSAPIIVRHNHSMFKRIPSTHETKQVW----- 282

QY 173 IKTSCTKEDCKDDILINCRRAATSVATCPIPPYVLAVGCDSSSVRIYDRMLGTRATGNY 232  
 Db 283 -SNGTSKSGEKDLQSYN-----VIAMGSDQRTITVW----- 312  
 QY 233 AGRTTGMVAR--FIPSHLNKSKRVTSCLCYSDGQILVSYSSDYLYL--FDPKDDTAR 288  
 Db 313 ----TTGS--ARPLFAVAKHFFGQS--VVDLSWSPDGLFACSLDGTVMATHFDPK----- 360  
 QY 289 ELATPSAERERELRQPPVKRLRLRGDWSOTGPRAPESERERDGEQSPNVSLMORMSDM 348  
 Db 361 ELGVRLTDTLDELKKSRYGDVR-----GQANLVESPAQLLLE----- 399  
 QY 349 LSRWFEASEVAQSNRGRSRPRGGTQSQSDISTLPTVPSPDLEVSETAMEVDTPAEQF 408  
 Db 400 -----FASTQAGSKRAASDVQOQVTKKSV-----SVETAKRRKSQVDDR 442  
 QY 409 LQSTSTMSQAQHSSTSPSTESPHSTPLSSPDSEQSQSVASGHHTHHQSDNNNEKLS 468  
 Db 443 NKAESTGOTLNKASTLNRYSSPVNQVYRRPDRKRILPEAVG---VPOENN----- 493  
 QY 469 KPCTGEPVLSLHYSTEGTITSTIKLNTDEWSSIASRIGSGHCKSEGOEESFVPOSSV 528  
 Db 494 -----IMIN-----GESHNFLPASAA 509  
 QY 529 OPPEGDSKTAPEESSEDVTKYQEGVSAENP-----VENHINITQSKDKFTAKPLDSS 581  
 Db 510 APAKGDGDPPEVSNRDLGKE--IVCRNPLKRSRITARATITES--LVIEKVPGTS 565  
 QY 582 GERNDLNLDRSCGVPEESAS 601  
 Db 566 GRDGLNVQSVGIKESST 585

## RESULT 23

D86731  
 hypothetical protein yihd [imported] - Lactococcus lactis subsp. lactis (strain IL140  
 C:Species: Lactococcus lactis subsp. lactis  
 C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
 C:Accession: D86731  
 R:Boilotin, A.; Winkler, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Eh  
 Genome Res. 11, 731-753, 2001  
 A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis  
 submitted to the Protein Sequence Database, April 2000  
 A:Reference number: A86625; MUID:21235186; PMID:11337471  
 A:Accession: D86731  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1063 <STO>  
 A:Cross-references: GB:AE005176; PID:g12723779; PIDN:AAK04950.1; GSPDB:GN00146  
 A:Experimental source: strain IL1403  
 C:Genetics:  
 A:Gene: yihd

Query Match 3.9%; Score 177.5; DB 2; Length 1063;  
 Best Local Similarity 24.3%; Pred. No. 0.064;  
 Matches 84; Conservative 46; Mismatches 163; Indels 53; Gaps 8;

QY 317 SDTGPRAPESEREDGEQSPNVSLMORMSDMLSRWFEEASEVAQSNRGRSRPRGGTS 376  
 Db 613 SQSSSSASSSSAAASSTSSSNVS-----SNTSSNSSEANTSSSTSNASSSSSSSEGS 667  
 QY 377 QSDISTLPTVPSPDLEVSETAMEVDTPAEQFLQPTSTSTMSQAQHSSTSPSTPHSTPL 436  
 Db 668 SSSNSSSSVASSSSVDSSQSSS-----AGVNSSSSSSAEGSSASSSSNESSVAS 717  
 QY 437 LSSPDSEQRQSVASGHHTHHQSDNNNEKLSPKPCTGEPVLSLHYSTEGTITSTIKLNT 496  
 Db 718 SSSVDSQSSAGVNG-----SSSSSESS-----ASSNSSEGSVASSSSVDSS 762  
 QY 497 DEWSS--IASSRGRIGHCKSEGOEESFVPOSSVQPPGDSKTAPEESSEDVTKYQEGV 554  
 Db 763 QSSSAGVNSSSSAEGSSASSSSNESSIASSSSVGSSQSSSTGVSSSSS----- 812  
 QY 555 SAENPVENHINITQSKDKFTAKPLDSSNGERNDLNLDRSCGVPEESASSEKAKEPETSDQT 614

```

QY 756 LLEADHNVCNQLQPHFPDFILASSGIDYDIKWTWSPLEESRIENRKLD-----EVT--RN 809
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      K L G : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 620 EKKLDSGTEKELVPLSTDTITINSSLGNEDSIYSLDDADASENLTDVPLMEIKTTPKY 679
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 810 ELMLEET-----RNTITVP-----ASFMLRLMASLN 835
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 680 EVISISGVSYSTSYENTVAMPQVEYITSPFMNDPFSNLN 719
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 25

B42680

nucleolus-cytoplasm shuttle phosphoprotein - rat

N;Alternate names: Nopp140 protein B

C;Species: Rattus norvegicus (Norway rat)

C;Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 17-Nov-2000

C;Accession: B43680; A42680; S27890; S27889; S30510; S30511

R;Meier, U.T.; Blobel, G.

Cell 70, 127-138, 1992

A;Title: Nopp140 shuttles on tracks between nucleolus and cytoplasm.

A;Reference number: A43680; MIM:93123542; PMID:1623516

A:Description: Nopp140 shuttles on tracks between nucleolus and cytoplasm.  
A:Reference number: S27889  
A:Accession: S27890  
A:Molecule type: mRNA  
A:Residues: 32-734 <ME2>  
A:Cross-references: EMBL:M94288; NID:g205751; PID:AAA1719.1; PID:g205752  
A:Accession: S27889  
A:Molecule type: mRNA  
A:Residues: 32-180, Q', 181-734 <ME12>  
A:Cross-references: EMBL:M94287; NID:g205749; PID:AAA1718.1; PID:g205750  
C:Superfamily: nucleolus-cytoplasm shuttle phosphoprotein  
C:Keywords: phosphoprotein

	Query Match	3.9%;	Score 176.5;	DB 2;	Length 734;
	Best Local Similarity	20.1%;	Pred. No. 0.045;		
	Matches 122;	Conservative 77;	Mismatches 260;	Indels 147;	Gaps 19;
QY	221	RMLGTRATGNVAGRGTCGWARFTPSHL-----NNKSCRVTSLCYSEDG---QEI	268		
	:       :	:       :	:       :		
Dd	21	RNSVVTRLTRSMADTG---LRRVVPDLYPLVLGLRDLNQLSEVASFAKATGATQDA	76		
	:       :	:       :	:       :		
QY	269	LVYSDDXYLLFPDKDDTARELKTPSAERREELQQPVKKRLRLRGWSDTGCPRARPESE	328		
	:       :	:       :	:       :		
Dd	77	NASSLLD-LYFSLWKSTKAPVKLOS-----NGPVAKKAKEITSSDSS---EDSS	123		
	:       :	:       :	:       :		
QY	329	RERDGEQSP-----NVLSMQMSDMLSRWFEEASEVAOSNRGRGRPR-----	372		
	:       :	:       :	:       :		
Dd	124	EEDKAAQVTFQKAAPAAPAKRASLPQHAGKAAKASAKASSESSSEEEEEKKKKPVQKAV	183		
	:       :	:       :	:       :		
QY	373	-----GGTSQSIDISTLPTVPSPDLEVSETAMEVDTPAE-----Q	407		
	:       :	:       :	:       :		
Dd	184	KPOKAVRPppPKAESSESSESDDDEAQTOKPAAATAAKAPTAKOTKAPKPGPPA	243		
	:       :	:       :	:       :		
QY	408	FLOP-----STSSWTSAQAHTSSTPESHSHTPL-----LSSPDE	443		
	:       :	:       :	:       :		
Dd	244	KAQPFKAANGKAGSSSSSSSSSSDDDEEEKKAAAPLUKTAPKQVVAKAPKVTAAPTQK	303		
	:       :	:       :	:       :		
QY	444	QRQSVASGHHTHHQSDNNNEKLSPKPGTGPVLSLRHYSTEGTTTTIKLNFTDEWSSTA	503		
	:       :	:       :	:       :		



[illegible]

RESULT 28  
T32271  
hypothetical protein ZC178.2 - *Caenorhabditis elegans*

Iga-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus influ  
C:Species: Haemophilus influenzae  
A:Variety: strain HK613  
C:Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 08-Dec-2000  
C:Accession: C41859  
R:Poulsen, K.; Reinholdt, J.; Kilian, M.  
J. Bacteriol. 174, 2913-2921, 1992  
A:Title: A comparative genetic study of serologically distinct Haemophilus influenzae  
A:Reference number: A41859; MUID:92234949; PMID:1373717  
A:Accession: C41859  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: nucleic acid  
A:Residues: 1-1849 <PDB>  
A:Experimental source: strain HK613  
A:Note: sequence extracted from NCBI backbone (NCBIP:97285)  
C:Superfamily: Iga-specific metalloendopeptidase  
C:Keywords: hydrolase; metalloproteinase

Query Match 3.9%; Score 175.5; DB 2; Length 1849;  
Best Local Similarity 21.8%; Pred. No. 0.17;  
Matches 106; Conservative 68; Mismatches 218; Indels 99; Gaps 19;  
QY 250 NTKSCRY-----TSLCYSDGQELVSYSDYIYLFDPKDDTARELKTSAEER 299  
DB 921 NTKSKVNVNKSATGNFTLQVADKTGP-----NHNETLTFDASNARNNLEVLANGSV 975  
QY 300 EELQPPVKRLRLRGDWS-----DTGPR--ARPESERERDEQSPNVLQRMDSMLSRW 352  
DB 976 D-----RGAWKYKLRNVNGRYDLYNVEKRNQTVDTNITTPNDI----- 1016  
QY 353 FEEASEVAQSNRGRSRPGTSGSDISTL--PTVPSSPDLEVSETAMEYDTPAEQFLQ 410  
DB 1017 -QADAPSAQSN-----NEEARVETPPVPAPAPATESAIAEQEPETRAETAQ 1062  
QY 411 P-----STSTMSAQAHSTSPSTESPHSTPLSPDSEQSQVSEASGHHTHQSDNNE 464  
DB 1063 PAMEETANSTETAPKSDTATQENPSE---SVPSSETTEKVAENPQENETVAKNEQE 1119  
QY 465 KLSKPGTGE-----PVLSLHYSTEGTTTITKLNFTDEWSSIASSSRGISGHCSEK 518  
DB 1120 ATEPTQNGEVAKEQDPVEANTQNEATQSEK---TEETQTAETKSEPTESVTSENQ 1176  
QY 519 EESFVQSSVQPPGDSKAPESSESDVTKYQGVSAENPVENHINITQSDKFTAK-PL 577  
DB 1177 PEKTVSQSTEDKVVVEKEKARVE--TEETQAPQVTSKEPP-----KQAEPAPEEPT 1228  
QY 578 DSNSEGRDLNDRSCGVPEESASSEKAKEPETSQDTESATNENNTNPEFQFTEATG 637  
DB 1239 DTNAEEAQAQQQTQTTVAEAETTPNSKPAETQOPS-----EKTNAEP-----VT 1275  
QY 638 PSAHEETSTRDSALQDITDSDDDPVL-IPGARYRAGPDRRSARVARIQEFFRRRKEREM 696  
DB 1276 PVVSENTATQPTETEETAKVEKTEQVVPQVASESPKQEQPA-AKPOAQTKPOAEPAPE 1334  
QY 697 EELDTLNTRRP 707  
DB 1335 NVLTTKNVGP 1345

RESULT 30  
T39903  
serine-rich protein - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Nov-2000  
C:Accession: T39903  
R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Rieger, M.  
submitted to the EMBL Data Library, November 1998  
A:Reference number: Z21889  
A:Accession: T39903  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-534 <LYN>  
A:Cross-references: EMBL:AL033534; PIDN:CAA22127.1; GSPDB:GN00067; SPDB:SPBC215.13

C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T32271  
R:Murray, J.; Wohldmann, P.; Beck, C.  
submitted to the EMBL Data Library, September 1997  
A:Description: The sequence of C. elegans cosmid ZC178.  
A:Reference number: Z21143  
A:Accession: T32271  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1459 <MUR>  
A:Cross-references: EMBL:AF024496; PIDN:AAE70340.1; GSPDB:GN00023; CBSP:ZC178.2  
A:Experimental source: strain Bristol N2; clone ZC178  
C:Genetics:  
A:Gene: CBSP:ZC178.2  
A:Map position: 5  
A:Introns: 1099/2

Query Match 3.9%; Score 176; DB 2; Length 1459;  
Best Local Similarity 21.4%; Pred. No. 0.12;  
Matches 143; Conservative 69; Mismatches 281; Indels 176; Gaps 31;  
QY 87 TTRSGHRANIFSAKLPCTNDQIVSCGQGVIFYT-----NVEQDAETNRQ 134  
DB 274 TTAGPSSSTATNSASSETPCNSETO---TSDGTSTMTVPNDSTTAGPSSSTATNSASSETP 329  
QY 135 QCFCHGVGTVEIMTVPNDDPTFLSCGEDGTVRFEDTRIKTCTKEDCKDDILLINCRRAA 194  
DB 330 CNSTQTSQDGTSTMTVPNDST---AGPSSSTAT-NSASSETPCNSETO---G 377  
QY 195 TSAVACPIPIYLLAVGSDSVRIYDRMLGLTRATGNVAGVGTGMVARFIPSHLNKSC 254  
DB 378 TSTMTVP-----NDST-----TAG-----PSSSTATNSA 400  
QY 255 RVTSICLYSE---DQQLVSYSDYIYLFDPKDDT-----ARELKTSAEER 300  
DB 401 SSETPCNSETO---PNDSTAGPSSSTATNSASSETPCNSETO--- 449  
QY 301 ELRQPPVKRLRLRGDWSDTGP-----RARPESERERDEQSPNVLQRMDSMLSRWE 354  
DB 450 --TSDGTSTMTVSSDSTTPGSSSTATNSASSETPCNSETO---SDGTFTMTVSSD--STTAG 505  
QY 355 BASEVAQSNRG---RGRSRPGTGSQSDISTLPTVPSSPDLEVSETAMEYDTPAEQFLQ 410  
DB 506 PSTVTVNSASSETPCNSETO---SDGTSTMTVSSDSTTPGSSSTATNSASSE--TPCNSETQ 563  
QY 411 PST--SSTMSAQAHSTSS--PTESPSTPLSPDSEQSQVSEASGHHTHQSDNNEKLSP 468  
DB 564 TSDGTSTMTVPNDSTTAGPSSSTATNSASSETPCNSETO---SDGTSTMT---VPNDSTTAGP 620  
QY 469 -----KPGTGEPLVSLHYSTEGTTTITKLNFTDEW--SSIASSSRGISGHCSEKQEE- 520  
DB 621 SSTVTVNSASSETPCNSETO---SDGTSTMTVSSDSTTAGPSSSTATNSASSETPCNSETO---SD 680  
QY 521 --SFVQSSVQPPGDSKAPESSE---DVTKYQGVSAENPVENHINITQSDKFTA 574  
DB 681 GTSTMTVSSDSTTAGPSSSTATNSASSETPCNSETO---SDGTSTMT-----VSSDSTA 732  
QY 575 KPLDSNSEGRDLNDRSCG-----VPDE-----SASSEKA--KE 607  
DB 733 GP---SSTATNSASSETPCNSETO---SDGTFTMTVPNDSTTAGPSSSTATNSASSETPCNSE 789  
QY 608 PPTSQDSTESATNENNTNPEFQFTEATGPS--AHEETSTRDSALQDITDSDDD--PVLI 664  
DB 790 TQTSQDGTSTMTVSSDST-----AGPSSSTATNSASSETPCNSETO---SDGTFTMTV 839  
QY 665 PGARYRAGP 673  
DB 840 PNDSTTAGP 848

RESULT 29  
C41859



QY 68 SGSDTKLVISNPYRKVLTIIRSHRANIFSAKFLP--CTNDKQIVSCSGDGVIFYINV 125  
DB 842 SGEDTT-----VVAVESGEGPASTSTSIPELSKDDQVTEASGETTAAAT 890  
QY 126 EDAETNRQCOFTCHYGTITIMTPNDPYTL-----SCGEDTVKWFTRIKTSCKED 181  
DB 891 EASEET-----TTSVATEGGETTVVAVESGEE-----PASSSTSIPTL 933  
QY 182 KDDILLINCRRAATVAICPPPIYLAVCCSSSVRIYDRMLCTRATGNYAGRTTGMV 241  
DB 934 SKDDQVTEASGETTAAATEAS---SETTTTAVTEGSEETTSVAVTEGSEET--- 987  
QY 242 ARFISHLNANKSCRTVSLCYSDGOBILVSYSD-----YIYLFDPKDDTARELKT 292  
DB 988 TSAVPEGENS---TTEAPAVTGSLEIPSSSESSSTTHDPISIPVITPKPVSSTIEN 1043  
QY 293 PSAEERREELRPPVKRLRLRGDWSGTGPAPRPERERDRGEOSPNVSLMQMSDMLSRW 352  
DB 1044 VMSKTSSEAAEKKIIGEHQTKGDDAG-----KEDENMFAFVTANPAGTSTTES 1094  
QY 353 FEAESEVAOSNRGRGRPRGGTSQSDISTPLTPVPSSPDLEVSETAMEVDTPAEQFLQPS 412  
DB 1095 AENVTTGTEDENIKAKELGKQFAADLAKL---AAKGVNLITETADAKDS-----GE 1144  
QY 413 TSTMSAQAHSTSSPTSPHSTPLLSPPSEQRQSVASGHHTHSDNNNEKLSPKPGT 472  
DB 1145 TAHVEDEQVSSSTESSIGSEETTTVNKETE-----HHEA-SGEEDDAPAFVT 1192  
QY 473 GEPVLSHYSTEGTITSTIKLN--FTDEWSSIA---SSSRGIGSHCKSGESESFVPOSS 527  
DB 1193 GAP-----TDSSTEASVTSITSAITDETSVAADSTSTASG-----EVQSSAIDSA 1240  
QY 528 VOPPEGDSKAP---EESSEDTV-----KYQEG---VSAENPVENHINI-TQ 568  
DB 1241 TVASEGTSSEATSVIESSGEEVTTDENLVTSTVAQLEGGSGITAEKDESDSVTTEAT 1300  
QY 569 SDRFTAKPLDSNGERNLMDRSCGVGPESASERAKETSDQ-----TSTESATNENN 624  
DB 1301 SQSTTVSESSDGSGE-----STVAPNDSETSTTESSTTDEGSGVTAESKDEESS 1352  
QY 625 TNPEPOTATGPSAHEETSTRDS---ALQDTDDSDDDPVLP 665  
DB 1353 TTEAPAVTSTSGSEDEEDSDPTHEFLGIDETMFKSLVP 1395  
RESULT 33  
A48819  
nuclear autoantigenic sperm protein - human  
C;Species: Homo sapiens (man)  
C;Date: 01-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 05-Nov-1999  
C;Accession: A48819  
R;O'Rand, M.G.; Richardson, R.T.; Zimmerman, L.J.; Widgren, E.E.  
Dev. Biol. 154, 37-44, 1992  
A;Title: Sequence and localization of human NASP: conservation of a Xenopus histone-binding domain  
A;Reference number: A48819; MUID:93050782; PMID:1426632  
A;Accession: A48819  
A;Status: preliminary  
A;Molecule type: nucleic acid  
A;Residues: 1-787 <OIR>  
A;Cross-references: GB:M97856; NID:g184432; PIDN:AAA36027.1; PID:g184433  
A;Experimental source: testis  
A;Note: sequence extracted from NCBI backbone (NCBIN:117050, NCBI:P117051)  
Query Match 3.9%; Score 174.5; DB 2; Length 787;  
Best Local Similarity 19.5%; Pred. No. 0.063;  
Matches 126; Conservative 96; Mismatches 228; Indels 195; Gaps 29;  
QY 283 KDDTARELKTSAEERREELRPPVKRLRLRGDWSGTGPAPRPERERDRGEO----- 335  
DB 124 EDESLEVNNDNIDEARELRLQVYDAMGEKEAKTKEDKSLAKETDKQDSEMEKGR 183  
QY 336 -----SPNVSLMQMSDMLSRWFEEASEVAOSNRGRGRPRGGTSQSDISTPLTPVSP 390

DB 184 EDMDISKSAEPQEKVDLTLDMLTETSEEA-----KGGAAPE-GPNEAEVTSKPEQVEVP 237  
QY 391 DLE-----VSETAMEVDTPAEQFLQPSSTTSSTMSAQAHSTSSPT 428  
DB 238 DAEKSVSGTDVQRECKRGQKGEQIVSLEEKPEVSEEQPVVILEKQCTAVEEA 297  
QY 429 ES--PHSTPLLSPPSEQRQSVASGHHTHSDNNNEKLSPKPGTGPVLSLHYSTEGT 486  
DB 298 ESLDPTVKPDVGGDPEEKVV-----TSENBAKAVLEQLVQGVPPPAEESPEVQ 348  
QY 487 TWTSTIKLNTDWSWSSSTASSRGIGSHCKSGESESFVPO--SSVQPPPE--GDSETKAPEES 543  
DB 349 TEAA-----EASAVEAGS---EYSEKPGQAPVLPKDGAVNGSVVGDQPTPIEPQTS 397  
QY 544 SEDVTYKQGVSAENPVENHINITSQDKFTAKPLDLSNGERNDLNDRSCGVPEESASSE 603  
DB 398 IERLTETKDGSGLEEKV-----RAKLVPSSQETKL 427  
QY 604 KAKEPETS-DOTSTESATNENNTNPEQTEATGPSAHEETSTRDSAL---QDTDDSD 659  
DB 428 SVESSEAAAGDVTQVQAQATEKSPEDKVOI-----AANEETQEREQMEGETEGSEE 482  
QY 660 DPVLIPGARYRAGPGRSARVARIQEFFRRRKRKEMEELDTLNI---RRPLVMYVKGH 716  
DB 483 DD-----KENDKTEEMPNDVLENKSLQENEEEGNLEAWMDLDAKIIFK-- 530  
QY 717 RNSRTMIKEANFGANFVMSGDCGHFIWDRHTAHLML---LEADNHV-----VNC 766  
DB 531 ---RQETKAQLAAQ-----AHLKIGESVSESENTVQVAAEFQSC 568  
QY 767 -----LOPHPPDPILASS---GIDYDIKIWSPLESRIFNKRLADEVITRNLML 815  
DB 569 LNLQEQVLEAH--DRLLAETHYQLGLAYG-----YNSQY-DEAAVQFSKSIEV 613  
QY 816 TRNTITVPASFMLRLASLHNRADRLEGDRSGSGOENENDEE 860  
DB 614 IENRMV-----LNE-----QVKEAEGS-SEYKKEIEE 640  
RESULT 34  
G86467  
hypothetical protein F7P12.4 [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Dec-2001  
C;Accession: G86467  
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon  
Chin, C.W.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A;Reference number: A86141; MUID:21016719; PMID:11130712  
A;Accession: G86467  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1587 <SFO>  
A;Cross-references: GB:AE005172; NID:g10092383; PIDN:AAG12790.1; GSPDB:GN00141  
C;Genetics:  
A;Map position: 1  
Query Match 3.9%; Score 174.5; DB 2; Length 1587;  
Best Local Similarity 20.0%; Pred. No. 0.16;  
Matches 191; Conservative 125; Mismatches 393; Indels 245; Gaps 47;  
QY 9 HLLWDPVRRKSLGLEDPDSRLRSYLRGRREFIQRLKLEATLNVDHGVNITCWNDTGEY--- 65  
DB 463 HL--DVTSEGTSSVPEDEDTYITTTMEISVPLASPNVLTARDSIKTKLOSTQDFQFP 520  
QY 66 ILGSGDDTKLVISNPYRKVLTIIRSG---HRANIFSAKFLPCTNDKQIVSCSGDGVIFY 122

Db 521 TMDLTVVAAAEKPPSSCVLGKSLGGVVEELGCFEL-FVAAEADK--CDVRGDSLHL 577  
QY 123 TNYEQDAETNRQOCFTCHYGTTEIMVNDPVTSLSCGEGIV-----RWFDI 171  
Db 578 NEISERMESS-----MSNKEDDP--FLAAKETSLPLSTDFINPETLWLYE 621  
QY 172 RIKTCTKEDCKDDILLINCRRATSVAIQPPPIPYILAVGCSDDSVRIYDRMLGTRATGN 231  
Db 622 DVQAS--PEFCYSSKANAENPSS-----CCSPSTEQI-----DC 655  
QY 232 YAGRTGTGMVARFIPFIPSHLNKSCRVTSLCYSDGQELLYSYSDYI-----YLDPP--K 283  
Db 656 FETSGCSAF-----DLAAEVESLHOE-----VSEETEFYKVMGVVSSEPLGK 700  
QY 284 DTARELTPSAERRE--ELRQPPVKRLRLR-----DTGPRARPESERE 330  
Db 701 ADIRSHENGESQEVSAEPVAKADIQSHENGEGTESGRQVIEVSPKSFSEAPT 760  
QY 331 RD-----GEOSPNNVSLMORMDML--SRWFEEASEVAQSNRGRSRPRGTT--S 376  
Db 761 LEILTEGAQGIIGSEFPSELAVETESNLLHQKSGIGETKNEIRSHEDYGETEDYGETCS 820  
QY 377 QSDISTLT--VPSPPDLEV-----SETAME-----VDTPAEQF 408  
Db 821 WPDIAVSPSSVSPPEPTLEILTDEARGLLGSEFLSEVTEIEMLLHQKSNVETKADIL 880  
QY 409 LQPTSTSTMSAQAHSTSPSPSTPSTPLSSPDSEQVSQVSEAGSHHTHQSD-----NNNE 464  
Db 881 IHEDYGETEVSQITIVSPNSFSAEPTLETEDSRQ--QARGLVGSDSEFQSEVAMKTECE 939  
QY 465 KLSKPKGTGEPVLSLHYSTEGTTSTIKNPTDEWSSSTASSRIGSHCKSEGESESVFP 524  
Db 940 NLLQKRNGETKVS--SRQASPVSDCLSTPKORLSI--NTDDIQSLC-----983  
QY 525 QSSVQPPGEGSETKAPESSEDVTKYQEGVSAEN--PVENHINITQDFTAKPLDSNS 581  
Db 984 -SSQPP--SESEVPATQD--QESGISETKPKTELLIGSGSEKYS--LSEIE 1034  
QY 582 GERN-DLNLDRSCGVPEESASSEKAKPEFTSDQSTSTESATNNTNPEPQFTEATGPSA 640  
Db 1035 GEENTDGLRSLRCPISALAAKTSSEDPKLEELSSDSSGSGEQNT-----PET 1082  
QY 641 HETSTSDALQTDSDDDPVLIPGARYRAGPGRDRSAVARAQEPPFRKKEKMEELD 700  
Db 1083 H--AVRDDVLCMDSSSTCNI-----WSRGRKAASVLKRTNKSQKQKQ--1125  
QY 701 TLNIRPLVKMYKGRNSRTMIKEANF--WGANFVMSGDCGHIFIDWRHTAEHLMLL 757  
Db 1126 --TGRQPKDKL--HRKQALSKSLSTIHHGAELP-----EFTPKEN--L 1168  
QY 758 EADNHVNCIQLPHPPDIPILASSGIDVDIKIWSPLESRIFNKLADVITRNLMLEPTR 817  
Db 1169 TPSSHMLKRLQDIG-DVKDSKSLKSLGKSCSLVHSSI--AVLASEATEPEIFTPKDE 1225  
QY 818 NITVPASFLMLASLNHADRLEGDRSGSGQ-----ENENED 858  
Db 1226 N--LTPSSHMLKRLREFGDIKDTK--GSSSKATRKPFEDIRMEENVYVQEPED 1275

## RESULT 35

A46194  
neurofilament protein NF-220, high-molecular-weight splice form - longfin squid  
C:Species: Loligo pealeii (longfin squid)  
C:Date: 22-Sep-1993 #sequence\_revision 25-Apr-1997 #text\_change 17-Mar-2000  
C:Accession: A46194  
R:Way, J.; Hellmich, M.R.; Jaffe, H.; Szaro, B.; Pant, H.C.; Gainer, H.; Battey, J.  
Proc. Natl. Acad. Sci. U.S.A. 89, 6963-6967, 1992  
A:Title: A high-molecular-weight squid neurofilament protein contains a lamin-like rod  
A:Reference number: A46194; MUID:92357751; PMID:1379729  
A:Accession: A46194  
A:Status: nucleic acid sequence not shown; not compared with conceptual translation  
A:Molecule type: mRNA

A:Residues: 1-1200 <WAY>  
A:Cross-references: GB:M94389; NID:g161291; PID:g161292  
A:Experimental source: stellate ganglion  
A>Note: sequence extracted from NCBI backbone (NCBIP:1113499)  
C:Superfamily: neurofilament triplet H protein  
C:Keywords: alternative splicing

Query Match 3.8%; Score 174; DB 2; Length 1200;  
Best Local Similarity 19.7%; Pred. No. 0.12;  
Matches 105; Conservative 79; Mismatches 220; Indels 130; Gaps 18;

QY 282 PKDDTARELTKTSAERREELRQPPVKRLRLRGDMSDTGPRAR-----324  
Db 682 PLSDT--IKSPVSE-----PAISPVSSVSGASGATSPAERTMSPIGSGSEKSAKSPV 731  
QY 325 -----PESERERDEQSPNVSLMORMDMLSRWFEEASEVAQSNRGRSR--RPRGCTTS 376  
Db 732 RSEATKSPVSEKSGSKSPVSEALSPVLSSEVSHSTAMSHTRSTASEKSVKSP 791  
QY 377 OSDISTLTPVSPDLEVSETAMEVDTPAEQFLOPSTS--STMSAQAHSTSS-----PTE 429  
Db 792 HSERTASPTAKSP-----IMEPAKSPKDESEKELSPERSEVMGSKSQITSSSAKSPVSE 848  
QY 430 SPHSTPLSSPDSEQVSQVSEAGSHHTHQSDNNNEKLSKPKGTGEPVLS--LHYSTEGTTT 488  
Db 849 KADSEKATSPTPEKVDDESSARSPTHSDRSEKQKASRPMTSDHIKSPIDFEKAESK 908  
QY 489 STIKLNFTE-----WSSIASSSRGIGSHCKSEGOEE-----SFVQSSVQVPEGDS 535  
Db 909 SALYSKSDHDKSPVPSEKAESEKARSVPVSEKAESEKARSVPVSEPAKSPVSEKAA 968  
QY 536 ETKAPEESSEDVTKYQEGVSAENPVENHINITQDFTAKPLDSNGE-----RNDLMLD 590  
Db 969 ESEKARSVPVSEKAESEKARSVPVPS--EKAESEKARSVPVSEKDESEKSAKSPSTDQ 1026  
QY 591 RSCGVPEESASSEKA-----KETSDQSTSTESA-----TNENTNPEPQFTE- 634  
Db 1027 ARSPVSEKAESEKARSVPVSEHVKSVPVSEKAESEKARSVPVSEHVKSVPVSEKAE 1086  
QY 635 -----ATGSAHEET-----STRDSALQDTD 655  
Db 1087 EKSASPVASEPAKSPVSEKAESEKSAKSPVSEHATSPVASEKSGSKSPAASEKA 1146  
QY 656 DSDDPVLIP-----GARYRAGPGRDRSAVARI-----QEFRRKKEKMEELD 700  
Db 1147 ESEKSPVSEKSGSKTAGSGSDGEGITTTTSSQERRSMSEKSKMSGSSD 1200

## RESULT 36

T34434  
hypothetical protein K06A9.1a - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 18-Feb-2000  
C:Accession: T34434  
R:Geisel, C.; Gattung, S.  
submitted to the EMBL Data Library, December 1996  
A:Description: The sequence of C. elegans cosmid K06A9.  
A:Reference number: Z21525  
A:Accession: T34434

A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-2232 <GE>

A:Cross-references: EMBL:U08046; PIDN:AAC70890.1; GSPDB:GN00028; CESP:K06A9.1a  
A:Experimental source: strain Bristol N2; clone K06A9  
C:Genetics:  
A:Gene: CESP:K06A9.1a  
A:Map position: X  
A:Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 20

Query Match 3.8%; Score 174; DB 2; Length 2232;  
Best Local Similarity 24.2%; Pred. No. 0.26;  
Matches 88; Conservative 37; Mismatches 188; Indels 50; Gaps 9;

[illegible]

```

RESULT 38
I38614
helicase II - human
C:Species: Homo sapiens (man)
C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 05-Nov-1999
C:Accession: I38614
R:Steyton, C.; Dabovic, B.; Gulisano, M.; Geacz, J.; Broccoli, V.; Giovanazzi, S.; B
numVol. Genet. 3 1957-1964, 1994
Article: Cloning and characterization of a new human Xql3 gene, encoding a putative h
I38614: MIMD:95179111: PMID:7874112

```

A:Molecule type: mRNA  
A:Residues: 1-1641 <RES>  
A:Cross-references: EMBL:U09820; NID:g606832; PTDN:AAC50069.1; PID:g606833  
C:Genetics:  
A:Gene: RAD54L

Db 542 RPRTRSAKAELEENQSRKQKRRRIKVKQEDSSSENKSNSEEEEEK -----EBEEE 595  
QY 646 TRDSALQDTDDDDPVLIPGARYRAGPDRRSARARI -----Q 684  
Db 596 EEEEEEEEDENDD-----SKSPGKGRKKIRKILKDDKLRLETQNALKEEERRK 646  
QY 685 EFFRRKKERKEEEL-----DTLNRPLVVMVYKGGHRSRTWIK 724  
Db 647 RIAERERERKRELVIEIEDASPTKPTIKLVLDDEDEETKEPLVQV-----HRMVIK 702  
QY 725 EANFWGANFVMSGDC 740  
Db 703 PHQVDGVQFMW---DC 715  
  
RESULT 39  
B98002  
Iga-specific metalloendopeptidase (EC 3.4.24.13) [imported] - Streptococcus pneumoniae.  
C:Species: Streptococcus pneumoniae  
C>Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 02-Nov-2001  
C:Accession: B98002  
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczyk, L.; Burgett, S.; DeHoff, B.S.; E  
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M  
Y, P.; Sun, P.M.; Winkler, M.E.  
J. Bacteriol. 183, 5709-5717, 2001  
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;  
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
A:Reference number: A97872; MUID:21429245; PMID:11544234  
A:Accession: B98002  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1963 <KUR>  
A:Cross-references: GB:AE007317; PIDN:AAK99846.1; PID:915458662; GSPDB:GN00174  
A:Gene: iga  
C:Superfamily: Streptococcus sanguis Iga-specific metalloendopeptidase  
C:Keywords: hydrolase; metalloproteinase  
  
Query Match 3.8%; Score 173; DB 2; Length 1963;  
Best Local Similarity 18.1%; Pred. No. 0.25;  
Matches 146; Conservative 120; Mismatches 299; Indels 242; Gaps 29;  
  
QY 111 IVSCSGDGVFTYTNVQDAETNRQCQFTCHYGTYYEI-----MTVPND--PYTFLSCGED 163  
Db 139 LVAGMNGVLVSVALEN-----HLLNNTDYELTSGEKLPLPKREISYTYIGYIKE 191  
QY 164 G-----TVNFFTRIKTSCTK----- 179  
Db 192 GKTTSDFEVSNOBKSAATPTKQKVDYNTVNFVDPHSTVQAIQEQTPVSSSTKPTQVQV 251  
QY 180 EDCDDILINCRRAATSAITCPPYPIYLVAGCSDSSVRIYDRMLGSDTGPRAPESEERERD 239  
Db 252 EKPFSDELINPREEKQ-----SSDSQBLAEHKNLETKKEEKISPKETG 297  
QY 240 MVARFTPLSHLNKSCRVTSLCTSDGQELVSYSS----- 274  
Db 298 VNTLNPQDELVSQMLKPKLRYEETIETKIDQEEIQENPDLAEGTVRVYKQEGKLGKV 357  
QY 275 DYIYLFD-PKDDTARELKTPSAEEERELRQPPVKRLRLRGDWSDTGPRARPESEERERD 333  
Db 358 EIVRFVSNKEVSEIVSTST-----TAPSPRIVEKGTGKTQVIRKQPETGVHEKD 409  
QY 334 EQS-----PNVSLMRMSDMLSRWFEEASEVAQSNRGRSRPRGTSQSDISTLPTV-- 386  
Db 410 VQSGAIVEPAIQ-----PELPEAVVSDKGEVEVQ-----TLPEAVV 446  
QY 387 -----PSPDLEVSETA-----MEVDTAPQFLQPTSTSSMSAQ 420  
Db 447 TDKGTEVQSPDPTVWSDKGEPEQVAPLPEYKGNIEQVKPETVPEKTRKQGEKTEEV- 505  
QY 421 ABSTSSPTESPHSTPLLSPPDSEQRQSVASGHHTH--HQSDNNNEKLSPKPGTGEPLVS 478  
Db 506 -----PVKPEETVPVNEGTEGTSIQEAPNPVQPAEESTTNSEKVS----- 549

QY 479 LHYTEGTTTTTIKLNFTDEWSSIASRRGIGSHCKSEGESEFVQSSVQPPGEGSETK 538  
Db 550 ---DTSSTTEVSSNPSTSTSVGENKPNHNDKSNENSEKT-VVEVPVNEGIVEGT 605  
QY 539 APESSSDVTYQGVSAENPVENHINITQSDKFTAKPLDLSNGERNDLNRDSCGVPEE 598  
Db 606 SNQETEPVQPAEE-----TQTN--SGKTANENTGEVSNKPSDKPPV-EE 648  
QY 599 SASSEK---AKEPETSQDTSTESATNE-----NNTNPEPQFOTATGSPSAHEETSTR 647  
Db 649 SNOPEKNGTATKPNSENGTTEPEKKLELNRVSDIELYSQTNGT-----YKQHVSL 704  
QY 648 DSALQDTD-----DSDDDPVLIPGARYRAGPDRRSA-----VARIQEFFRRR- 690  
Db 705 DGIPENTDTYFVKVKSFAKVDYIPVASITEEKRNGQSVYKITAKAEKLQLEKNKYVDN 764  
QY 691 ----KERKEEELDTLNRRLPLVMVYKGGHRSRTMIKEANFWGANFVMSGDCGHIFTW 746  
Db 765 FTFVLDKKAKEENTNFTSFSNLVKAINQNPSTGYTHLAASLN---ANEVELGPD-ERSYIK 820  
QY 747 DRHTAEHLMLLEADNHNVCLOPHFED 773  
Db 821 DTFTRLIGEKGDKGKNYAIYNLKKPLFE 847  
  
RESULT 40  
T30989  
serine/threonine protein kinase NIK - mouse  
N:Alternate names: Nck interacting kinase  
C:Species: Mus musculus (house mouse)  
C>Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
C:Accession: T30989  
R:Su, Y.-C.; Han, J.; Xu, S.; Cobb, M.; Skolnik, E.Y.  
EMBO J. 16, 1279-1290, 1997  
A:Title: NIK is a new Ste20-related kinase that binds NCK and MEK1 and activates the  
A:Reference number: Z20954; MUID:97280817; PMID:9135144  
A:Accession: T30989  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1233 <NUY>  
A:Cross-references: EMBL:U88984; NID:g1872545; PID:g1872546; PIDN:AAC53165.1  
C:Keywords: protein kinase  
  
Query Match 3.8%; Score 172.5; DB 2; Length 1233;  
Best Local Similarity 20.8%; Pred. No. 0.14;  
Matches 139; Conservative 75; Mismatches 261; Indels 193; Gaps 29;  
  
QY 283 KDDTARELKTPSAEEERELRQPPVKRLRLRGDWSDTGPRARPESEERERDGEQSPNVSLM 342  
Db 409 EQQRREAREARQREQRREQREKR-RL-----EELERKEEERERAEERKRV--- 460  
QY 343 ORMSDMLSRWFEEAS---EVAQS-----NRGRSRPRGGTSGSDISTLPTV 386  
Db 461 EREQYIIRQLREEQRHLEILQQLLOEQAMLLHRRHPHAQQPPPPQQQDRSKPSFHA 520  
QY 387 PS-SPDLEVSETAMEV-----DTPAEQFLQ-----PSTS 414  
Db 521 PEKPHYDPADRARREVQMSHLASLNKNVSPVSRSHSFSDPSKFAHHHLRSQDPCPPSR 580  
QY 415 STMSAQAHSTS---SPT-----ESPSTPLLS---SPDSEQRQSVASGHHTHQ 458  
Db 581 EGLSQSDSKSEVPEPTQKAWRSRSDSEVPVPVTRTTSRSPVLSRRDSPLOGGGQNSQ 640  
QY 459 SDNNN-----EKLSPKPGTGEVLSLHYSTEGTTTSTIKLNTFDEWSSIAS 504  
Db 641 AQQRNTSSIEPRLWVERKLVPRPGSG-----SSSGS 674  
QY 505 SSRGI---GSHCKSE-GQSEFVQSSVQ-----PPEGDSETKAPESS- 545  
Db 675 SNSGSGQPSGHPGSGGGERFRVRSRSGSEKSPSPQSAKKPDDKKEVFRSLKPAGEV 734  
QY 546 DVTYQGVSAENPVENHINIT--QSDKFTAKPLDLSNGERNDLNRDSCGVPEE--SAS 601

Db 735 DLTALAKELRAVEDVRPPHKVTDYSSSSSESGTTDEEEDVEQEGADDSTSGPEDTRAAS 794  
QY 602 SEKAKEPETSDOTSTESATNENNTNPEQFQTEAUGPSAHEETSTRDSALQDTHDDSD-- 659  
Db 795 SPNLSNGETE---SVKTMIVHDDVESEPAWTPSKEGTLIVRQTSASSTLQKHKSST 851  
QY 660 ----DPVLI-----PGARYRAGPCD--RRSAVARIQEFFRERKERKEM 696  
Db 852 PFIDPRLLQISPSSGTTVTSVVGFSCDGLRPEAIRQDPTKGSVVNVNPTNTR----- 904  
QY 697 EELDTLNIIRPLVKMYKGRHNSRTMIKEANFWGVMSGDCGHIFWDRHTAHLML 756  
Db 905 PQSDTPEIRK-----YKKFENSEILC--AALWGVNLLV-GTESG-----LML 943  
QY 757 LE--ADNHVVNCLQPHFPDPIILASSGIDYDIKIWSPLEESRIFNRKLADEVITRNELMLE 814  
Db 944 LDRSGQGVVPLISRRRFQOMVLEGLNVLVTISGKKDKLRVYLSWLRNKLHNDPEVE 1003  
QY 815 ETRNTITV 822  
Db 1004 KQGWTTV 1011

Search completed: March 18, 2003, 15:36:23  
Job time : 76 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: March 25, 2003, 20:18:38 ; Search time 7524 Seconds  
(without alignments)  
9979.431 Million cell updates/sec

Title: US-09-781-693A-1  
Perfect score: 2580  
Sequence: 1 atgtctcgggggtggtctccta.....atgaataatgagatgagaa 2580

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 120 summaries

Database :

GenEmbl:\*

- 1: gb\_ba:\*
- 2: gb\_hgt:\*
- 3: gb\_in:\*
- 4: gb\_cm:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sv:\*
- 13: gb\_un:\*
- 14: gb\_vt:\*
- 15: em\_ba:\*
- 16: em\_fun:\*
- 17: em\_hum:\*
- 18: em\_in:\*
- 19: em\_mu:\*
- 20: em\_cm:\*
- 21: em\_or:\*
- 22: em\_ov:\*
- 23: em\_pat:\*
- 24: em\_ph:\*
- 25: em\_pl:\*
- 26: em\_ro:\*
- 27: em\_sts:\*
- 28: em\_un:\*
- 29: em\_vt:\*
- 30: em\_hgt\_hum:\*
- 31: em\_hgt\_inv:\*
- 32: em\_hgt\_other:\*
- 33: em\_hgt\_mus:\*
- 34: em\_hgt\_pln:\*
- 35: em\_hgt\_rod:\*
- 36: em\_hgt\_nam:\*
- 37: em\_hgt\_vrt:\*
- 38: em\_sv:\*
- 39: em\_hgt\_hum:\*
- 40: em\_hgt\_mus:\*
- 41: em\_hgtg\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2573.6	99.8	3189	6	AX076993	AX076993 Sequence
2	2573.6	99.8	3189	6	AX113966	AX113966 Sequence
3	2508.4	97.2	3022	6	AX078303	AX078303 Sequence
4	2332.6	90.4	3420	6	AX076995	AX076995 Sequence
5	2332.6	90.4	3420	6	AX113968	AX113968 Sequence
6	2140	82.9	2448	6	AX060956	AX060956 Sequence
7	2140	82.9	2448	6	AX060956	AX060956 Sequence
8	2076.4	80.5	2325	9	AF124434	AF124434 Homo sapi
9	2073	80.3	2727	9	AK074378	AK074378 Homo sapi
10	1594.4	61.8	2024	9	BC025262	BC025262 Homo sapi
11	1446.4	56.1	1952	9	BC025262	BC025262 Homo sapi
12	1378.4	53.4	2803	9	AK061965	AK061965 Homo sapi
13	1287.2	49.9	162314	9	AK061965	AK061965 Homo sapi
14	1285.6	49.8	159098	2	AC016484	AC016484 Homo sapi
15	1067.4	41.4	1726	9	AK000984	AK000984 Homo sapi
16	567.4	22.0	65517	2	AC107981	AC107981 Homo sapi
17	328.8	12.7	48438	9	HS702319	HS702319 Homo sapi
18	328.8	12.7	163230	2	AC026140	AC026140 Homo sapi
19	266.6	10.3	74729	9	HS703H14	HS703H14 Human DNA
20	242.2	9.4	251	11	G05536	G05536 human STS
21	217.4	8.4	158324	2	AC069147	AC069147 Homo sapi
22	217.4	8.4	163230	2	AC026140	AC026140 Homo sapi
23	202.8	7.9	168307	2	AC130131	AC130131 Rattus no
24	182.6	7.1	179269	2	AC116374	AC116374 Mus muscu
25	166.6	6.5	100694	2	AC127803	AC127803 Rattus no
26	166.6	6.5	168307	2	AC130131	AC130131 Rattus no
27	159.6	6.2	179269	2	AC116374	AC116374 Mus muscu
28	142.4	5.5	100694	2	AC127803	AC127803 Rattus no
29	125.2	4.9	59143	2	AC100993	AC100993 Mus muscu
30	125.2	4.9	59143	2	AL732543	AL732543 Mus muscu
31	98	3.8	158324	2	AC069147	AC069147 Homo sapi
32	96.4	3.7	102228	9	HS295C6	HS295C6 Human DNA
33	76	2.9	1765	8	AY120722	AY120722 Arabidops
34	71.2	2.8	2594	8	AY039964	AY039964 Arabidops
35	70.6	2.7	2034	6	AX370488	AX370488 Sequence
36	70.6	2.7	2966	9	AK023101	AK023101 Homo sapi
37	70.6	2.7	3267	9	AK023778	AK023778 Homo sapi
38	70.6	2.7	4211	6	AX370510	AX370510 Sequence
39	70.6	2.7	4305	9	BC032523	BC032523 Homo sapi
40	69.4	2.7	798	9	HS295C061	HS295C061 Sequence
41	66.8	2.6	84499	8	ATT12J5	ATT12J5 Homo sapi
42	66.8	2.6	195165	8	ATCHRIV82	ATCHRIV82 Arabidops
43	66.6	2.6	114897	2	AP003624	AP003624 Oryza sat
44	66.6	2.6	141534	2	AP003685	AP003685 Oryza sat
45	65	2.5	110980	8	ATF9K21	ATF9K21 Arabidops
46	64.2	2.5	2834	9	AK098114	AK098114 Homo sapi
47	63.8	2.5	1887	6	AX370484	AX370484 Sequence
48	63.8	2.5	2206	6	AX370508	AX370508 Sequence
49	63.8	2.5	2286	3	AC004296	AC004296 Drosophil
50	63.8	2.5	78054	3	AC004296	AC004296 Drosophil
51	63.8	2.5	120625	3	AC017563	AC017563 Drosophil
52	63.8	2.5	136164	3	AC099006	AC099006 Drosophil
53	63.8	2.5	162492	3	AC099039	AC099039 Drosophil
54	63.8	2.5	255386	3	AE003800	AE003800 Drosophil
55	63.4	2.5	81631	9	HS745114	HS745114 Human DNA
56	63	2.4	81740	9	AL663123	AL663123 Human DNA
57	61	2.4	1731	6	AX370534	AX370534 Sequence
58	61	2.4	5091	9	AB028960	AB028960 Homo sapi
59	60.8	2.4	1596	8	AY084990	AY084990 Arabidops
60	60.8	2.4	91938	8	ATF20M13	ATF20M13 Arabidops
61	60.8	2.4	93760	8	ATF22M13	ATF22M13 Arabidops
62	60.8	2.4	107740	8	AC069558	AC069558 Genomic S
63	60.8	2.4	199789	8	ATCHRIV89	ATCHRIV89 Arabidops
64	60.2	2.3	773	6	AX187045	AX187045 Sequence
65	60	2.3	773	8	AY086855	AY086855 Arabidops



Db 716 GATTGTAAGAGATATTTAACTGTCAGCGTGTGCCAGTCTGTGCTATTGCG 775  
QY 601 CCACCAATACCAATATACCTTGTGCTGTTGTTCTCAGACCTCAGTACGAATATGAT 660  
Db 776 CCACCAATACCAATATACCTTGTGCTGTTGTTCTCAGACCTCAGTACGAATATGAT 835  
QY 661 CGCGAATGCTGGCACAAGAGCTACAGGAAATATCAGGCTCAGGAGTACTGGATG 720  
Db 836 CGCGAATGCTGGCACAAGAGCTACAGGAAATATCAGGCTCAGGAGTACTGGATG 895  
QY 721 GTTGCCCGTTTATTCCTCCCACTTAAATAAGTCCCTGCGAGTGCATCTCTGTGT 780  
Db 896 GTTGCCCGTTTATTCCTCCCACTTAAATAAGTCCCTGCGAGTGCATCTCTGTGT 955  
QY 781 TACAGTGAAGATGGTCAAGAGATCTCGTGTAGTACTCTTCAGATTACATATATCTTTT 840  
Db 956 TACAGTGAAGATGGTCAAGAGATCTCGTGTAGTACTCTTCAGATTACATATATCTTTT 1015  
QY 841 GACCCGAAAGATGATACAGCAGAGAACTTAAACCTCTCTGCGGAAGAGAGAGAA 900  
Db 1016 GACCCGAAAGATGATACAGCAGAGAACTTAAACCTCTCTGCGGAAGAGAGAGAA 1075  
QY 901 GAGTTGCGACACACCGAGTTAAGCGTTTGAGACTTCGTGTGATTGGTCAATCTGGA 960  
Db 1076 GAGTTGCGACACACCGAGTTAAGCGTTTGAGACTTCGTGTGATTGGTCAATCTGGA 1135  
QY 961 CCCAGCAAGCGCGAGAGTGAACGAGACGAGATGAGAGCAGAGTCCCAATGTGCA 1020  
Db 1136 CCCAGCAAGCGCGAGAGTGAACGAGAGAGATGAGAGCAGAGTCCCAATGTGCA 1195  
QY 1021 TTGATGAGAGATGTCGTGATATGTTATCAAGATGGTTGAAGAGCAGTGAATTT 1140  
Db 1196 TTGATGAGAGATGTCGTGATATGTTATCAAGATGGTTGAAGAGCAGTGAATTT 1255  
QY 1081 CAAAGCAATAGAGCAGAGAGATCTCGACCCAGAGAGTGAACAGTCAATCAGATAT 1200  
Db 1256 CAAAGCAATAGAGCAGAGAGATCTCGACCCAGAGAGTGAACAGTCAATCAGATAT 1315  
QY 1141 TCAACTCTTCTACGTCCTCAAGTCCCTGATTGGAAGTGAAGTGAACAGTCAATCAGATAT 1260  
Db 1316 TCAACTCTTCTACGTCCTCAAGTCCCTGATTGGAAGTGAAGTGAACAGTCAATCAGATAT 1375  
QY 1201 GTAGATCTCCAGTGAACATTTCTCAGCCTCTGAGCATCTGGACACACATCATCTGTAT 1260  
Db 1376 GTAGATCTCCAGTGAACATTTCTCAGCCTCTGAGCATCTGGACACACATCATCTGTAT 1335  
QY 1261 GCTATTCGACATCATCTCCACAGAAAGCCCTCATCTACTCTTGTCTATCTTCCA 1320  
Db 1436 GCTATTCGACATCATCTCCACAGAAAGCCCTCATCTACTCTTGTCTATCTTCCA 1495  
QY 1321 GACAGTGAACAAAGCAGTCTGTTGAGGCATCTGGACACACATCATCTGTAT 1380  
Db 1496 GATAGTGAACAAAGCAGTCTGTTGAGGCATCTGGACACACATCATCTGTAT 1555  
QY 1381 ACAATAATGAAAGTGAAGCCCAACAGGACAGTGAACAGTGAACAGTGAACAGTGAAC 1440  
Db 1556 ACAATAATGAAAGTGAAGCCCAACAGGACAGTGAACAGTGAACAGTGAACAGTGAAC 1615  
QY 1441 TACAGCAGAGAGGACACTACAGCACAATAAAGTGAACAGTGAACAGTGAACAGTGAAC 1500  
Db 1616 TACAGCAGAGAGGACACTACAGCACAATAAAGTGAACAGTGAACAGTGAACAGTGAAC 1675  
QY 1501 AGTATAGCATCAAGTCTAGAGGAATGGGAGCCATTCGAATCTGAGGTCAGGAGGAA 1560  
Db 1676 AGTATAGCATCAAGTCTAGAGGAATGGGAGCCATTCGAATCTGAGGTCAGGAGGAA 1735  
QY 1561 TCTTTTCGTCACAGCTCAGTGAACACAGAGAGAGTGAACAGTGAACAGTGAACAGTGAAC 1620  
Db 1736 TCTTTTCGTCACAGCTCAGTGAACACAGAGAGAGTGAACAGTGAACAGTGAACAGTGAAC 1795  
QY 1621 GAAGAAATCATCAGAGGATGTCAGAAAATATCAGGAGAGATCTGCGAGAAACCCAGTT 1680  
Db 1796 GAAGAAATCATCAGAGGATGTCAGAAAATATCAGGAGAGATCTGCGAGAAACCCAGTT 1855

QY 1681 GAGAACCATATCAATATAACACAATCAGATAGTTTACAGCCCAAGCCATTGGATTCCAAAC 1740  
Db 1856 GAGAACCATATCAATATAACACAATCAGATAGTTTACAGCCCAAGCCATTGGATTCCAAAC 1915  
QY 1741 TCAGGAGAAAGAAATGACCTCAATCTTGTATCGCTCTTGTGGGTTCCAGAGAAATCTGCT 1800  
Db 1916 TCAGGAGAAAGAAATGACCTCAATCTTGTATCGCTCTTGTGGGTTCCAGAGAAATCTGCT 1975  
QY 1801 TCATCTCAAAAGAAAGCAAGGACCAAGAACTTCAGATCAGACTAGCAGTGCATACC 1860  
Db 1976 TCATCTCAAAAGAAAGCAAGGACCAAGAACTTCAGATCAGACTAGCAGTGCATACC 2035  
QY 1861 AATGAAATTAACACCAATCTGAGCCCTCAGTTCACAAACAGAAAGCCACTGGGCTTCAGCT 1920  
Db 2036 AATGAAATTAACACCAATCTGAGCCCTCAGTTCACAAACAGAAAGCCACTGGGCTTCAGCT 2095  
QY 1921 CATGAAAGAAATCCACCAAGGAGTCTGCTCTTCAGGACACAGATCAGACTGATGATGAC 1980  
Db 2096 CATGAAAGAAATCCACCAAGGAGTCTGCTCTTCAGGACACAGATCAGACTGATGATGAC 2155  
QY 1981 CCAGTCTCTGATCCAGGTCGCAAGGTATCGAGCAGGACCTGGTGTATGATGATGATGAT 2040  
Db 2156 CCAGTCTCTGATCCAGGTCGCAAGGTATCGAGCAGGACCTGGTGTATGATGATGATGATGAT 2215  
QY 2041 GCCCGTATTCAGAGTCTTCAGACGAGAAAGAAAGAAAGAAATGGAATTTGGAT 2100  
Db 2216 GCCCGTATTCAGAGTCTTCAGACGAGAAAGAAAGAAAGAAATGGAATTTGGAT 2275  
QY 2101 ACTTTGAAACATFAGAGCGCGTCTAGTAAAATGTTTATAAGGCCATCGCAACTCCAGG 2160  
Db 2276 ACTTTGAAACATFAGAGCGCGTCTAGTAAAATGTTTATAAGGCCATCGCAACTCCAGG 2335  
QY 2161 ACAATGATAAAGAGCCCAATTTCTGGGTGCTAACTTTGTAATCAGTGTCTGACTGT 2220  
Db 2336 ACAATGATAAAGAGCCCAATTTCTGGGTGCTAACTTTGTAATCAGTGTCTGACTGT 2395  
QY 2221 GCCCACATTTTCATCTGGGATCGGCACACTGCTGAGCAATTTGATGCTTCTGGAAGCTGAT 2280  
Db 2396 GCCCACATTTTCATCTGGGATCGGCACACTGCTGAGCAATTTGATGCTTCTGGAAGCTGAT 2455  
QY 2281 AATCATGTGGTAAACTGCCCTGACCCACATCGCTTGGACCAATTTTACCCCTCATCTGCG 2340  
Db 2456 AATCATGTGGTAAACTGCCCTGACCCACATCGCTTGGACCAATTTTACCCCTCATCTGCG 2515  
QY 2341 ATAGATTATGACATAAAGATCTGCTCACCATTAGAGAGTCAAGGATTTTAAACCGGAAA 2400  
Db 2516 ATAGATTATGACATAAAGATCTGCTCACCATTAGAGAGTCAAGGATTTTAAACCGGAAA 2575  
QY 2401 CTTGCTGATGAAGTTATTAACCTCGAAACGAACTCATGCTGGAAGAACTAGAAACACCAT 2460  
Db 2576 CTTGCTGATGAAGTTATTAACCTCGAAACGAACTCATGCTGGAAGAACTAGAAACACCAT 2635  
QY 2461 ACAGTTCCAGCCCTTCATGTTGAGGATGTTGGTTCCTTAATCATATCCGAGCTGAC 2520  
Db 2636 ACAGTTCCAGCCCTTCATGTTGAGGATGTTGGTTCCTTAATCATATCCGAGCTGAC 2695  
QY 2521 CGTTTGGAGGTCAGATCAGAAAGCTCTGCTCAAGAGAAATGAAATGAGGATGAGGAA 2580  
Db 2696 CGTTTGGAGGTCAGATCAGAAAGCTCTGCTCAAGAGAAATGAAATGAGGATGAGGAA 2755

## RESULT 2

AX113966

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AX113966  
Sequence 1 from Patent EP1106690.

AX113966.1 GI:13940140

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

3189 bp  
DNA  
linear  
PAT 01-MAY-2001

REFERENCE 1 (bases 1 to 3189)  
AUTHORS el Shami,A.S., Menon,S.N. and French,C.K.  
TITLE Polynucleotide encoding autotransmembrane proteins associated with endometriosis  
JOURNAL Patent: EP 1106690-A 1 13-JUN-2001;  
DIAGNOSTIC PRODUCTS CORPORATION (US)  
FEATURES  
source Location/Qualifiers  
1. .3189  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
176. .2758  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC37867.1"  
/db\_xref="GI:13940141"  
translation="MSRGGSYPHLLMDVRKRLFLGLEDPSRLSRRLRYLGRBFRTORLKL  
ATLVHDCVNTICWNTDGYILSGSDTKLVISNYSYRKVLTIIISGRHRANFSAK  
LPCTNDQIVSCSDQVIFVTNVEQDAETNRQCTCHYGTTEIIMVNDPFTFLSC  
GEDTVRFWTRIKTSCTKEDCKDDILINCRRAATVAICPPPIYLAQSDSSVRI  
YDRMLCTVRATGAGTGTGMVAREFTPSHLNKKSCRTSLCYSEDEQELLYSSDY  
IYLFDPKDDTARELKTSAERREELRQPPVKRLRLGWDSDTGPRAPESEERDGE  
QSNVSLMQRMSLRSRFRSEASVQAQSGRSRPRGCTSDSDSTLPTVSPDOL  
EVSTAMEVDTPAEQFLOPSTSTMSQAHSSTSPESPHSTPLISDSEQSQSVEA  
SGHTHQSDNNNEKLSKPGTGEPLVLSLHYSTBTGTTSTIKLNFDEWSSIASSE  
IGSHCKSQSDSESVQSSQPPQPEGDSKAPESSEKDEKPESTQSTSTESATNEN  
TQSDKFTAKPLDSNGRNLNDRSCGVPESEASSEKAKEPESTQSTSTESATNEN  
TNPEPQFTEATGSAHETSTRDSDLDSDPVLIPGARYRAGPDGRSAVAR  
IQEPRFRKEREKMEELDTLIRPLVMYVKGHRNSRTMIKEANFAGNFMVFTGSC  
GHIFLWDRHTAHEMLLEADNHNVCLOPHFDPILASSGIDYDIKWTSPLEESRIFN  
RKLDEVTIRNELMLEETRNITVPASEFMLRLASLNHIRADRLEGRSESGGENEN  
EDEE"

BASE COUNT 977 a 661 c 724 g 827 t

ORIGIN

Query Match 99.8%; Score 2573.6; DB 6; Length 3189;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 2576; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATGCTCGGGGTGGCTCCTACCCACACCTGTTGTGGGAGCTGAGGAAAGTCCCTCGG 60  
DB 176 ATGCTCGGGGTGGCTCCTACCCACACCTGTTGTGGGAGCTGAGGAAAGTCCCTCGG 235  
QY 61 CTGGAGGACCCGTCGGGCTCGGAGTGCCTACCTCGGAGAGAGAAATTTATCCAAAGA 120  
DB 236 CTGGAGGACCCGTCGGGCTCGGAGTGCCTACCTCGGAGAGAGAAATTTATCCAAAGA 295  
QY 121 TTAAGCTTGAAGCAACCCCTTAATGTCATGATGTTGTTGTTATACATCTCTGGAAT 180  
DB 296 TTAAGCTTGAAGCAACCCCTTAATGTCATGATGTTGTTGTTATACATCTCTGGAAT 355  
QY 181 GACACTGGAGAAATATATTTATCTGGCTCAGATGACACCAATTTAGTAAATAGTAACTCT 240  
DB 356 GACACTGGAGAAATATATTTATCTGGCTCAGATGACACCAATTTAGTAAATAGTAACTCT 415  
QY 241 TACAGCAAGAAAGTTTGAACAATTTCTGTCAGGCGACCGAGCAACATATTTAGTGCA 300  
DB 416 TACAGCAAGAAAGTTTGAACAATTTCTGTCAGGCGACCGAGCAACATATTTAGTGCA 475  
QY 301 AAGTTCTTACCTGTACAAATGATAACAGATTTGTTATCTGCTCGAGATGGAGTAATA 360  
DB 476 AAGTTCTTACCTGTACAAATGATAACAGATTTGTTATCTGCTCGAGATGGAGTAATA 535  
QY 361 TTTTATACCAAGTTGAGCAAGATGCGAAACCAACAGACAAATGCCAATTTACGTGTCAT 420  
DB 536 TTTTATACCAAGTTGAGCAAGATGCGAAACCAACAGACAAATGCCAATTTACGTGTCAT 595  
QY 421 TATGGAATCTATTGAGATTTAGCTGTACCGAATGACCCCTTACCTTTCTCTCTGT 480  
DB 596 TATGGAATCTATTGAGATTTAGCTGTACCGAATGACCCCTTACCTTTCTCTCTGT 655  
QY 481 GGTGAAGATGGAACCTGTTAGTGGTTTTCATACAGCATCAAAACCTAGCTGCACAAAAGAA 540  
DB 656 GGTGAAGATGGAACCTGTTAGTGGTTTTCATACAGCATCAAAACCTAGCTGCACAAAAGAA 715

QY 541 GATTCGTAAGATGATATTTTAACTAGTGTGACGCTGTGCGACGCTGCTGCTGCTATTTC 600  
DB 716 GATTCGTAAGATGATATTTTAACTAGTGTGACGCTGTGCGACGCTGCTGCTGCTATTTC 775  
QY 601 CCACCAATACCATTATACCTTCTGTTGTTCTGACAGCTAGTACGAAATATATGAT 660  
DB 776 CCACCAATACCATTATACCTTCTGTTGTTCTGACAGCTAGTACGAAATATATGAT 835  
QY 661 CGCGGAATCTGCGCAAGAGCTACAGGGAATATATGACGTCGAGGAGGACTACTGGAATG 720  
DB 836 CGCGGAATCTGCGCAAGAGCTACAGGGAATATATGACGTCGAGGAGGACTACTGGAATG 895  
QY 721 GTTGCCCGTTTATTCCTTCCCATCTTATTAATAAGTCTGACAGTGCATCTCTGTGT 780  
DB 896 GTTGCCCGTTTATTCCTTCCCATCTTATTAATAAGTCTGACAGTGCATCTCTGTGT 955  
QY 781 TACAGTGAAGATGGTCAAGAGATCTCTGTTAGTACTCTTCAGATTACATATATCTTTT 840  
DB 956 TACAGTGAAGATGGTCAAGAGATCTCTGTTAGTACTCTTCAGATTACATATATCTTTT 1015  
QY 841 GACCCGAAAGATGATACAGCAGCAGAACTTAAACTCTCTGCGAAGAGAGAGAGAA 900  
DB 1016 GACCCGAAAGATGATACAGCAGCAGAACTTAAACTCTCTGCGAAGAGAGAGAGAA 1075  
QY 901 GAGTTGCGCAACACCACAGTTAAAGCGTTTGAAGACTTTCGTTGGTGAATGGTCAGATACTGA 960  
DB 1076 GAGTTGCGCAACACCACAGTTAAAGCGTTTGAAGACTTTCGTTGGTGAATGGTCAGATACTGA 1135  
QY 961 CCCAGAGCAAGCGCGGAGAGATGACAGAGACAGATGGAGAGCAGAGTCCCAATGTGCA 1020  
DB 1136 CCCAGAGCAAGCGCGGAGAGATGACAGAGACAGATGGAGAGCAGAGTCCCAATGTGCA 1195  
QY 1021 TTGATGACAGAGATGCTGATATGTTATCAAGATGGTTTGAAGAAGCAAGTGAAGTTGCA 1080  
DB 1196 TTGATGACAGAGATGCTGATATGTTATCAAGATGGTTTGAAGAAGCAAGTGAAGTTGCA 1255  
QY 1081 CAAAGCAATAGAGCAGGAGAGATCTCGACCCAGAGGTGGAACAGTCAATCAGATATT 1140  
DB 1256 CAAAGCAATAGAGCAGGAGAGATCTCGACCCAGAGGTGGAACAGTCAATCAGATATT 1315  
QY 1141 TCACCTCTCTACGCTCCCATCAAGTCTCTGATTTGGAAGTGAAGTGAACCTGCAATGAA 1200  
DB 1316 TCACCTCTCTACGCTCCCATCAAGTCTCTGATTTGGAAGTGAAGTGAACCTGCAATGAA 1375  
QY 1201 GTAGATCTCCAGCTGGAACAATTTCTTCAGCCTTCTACATCTCTACATGTCTAGCTCAG 1260  
DB 1376 GTAGATCTCCAGCTGGAACAATTTCTTCAGCCTTCTACATCTCTACATGTCTAGCTCAG 1435  
QY 1261 GCTCATTCGACATCATCTCCACAGAAAGCCCTCATTTCTCTCTTCTGCTATCTTCTCCA 1320  
DB 1436 GCTCATTCGACATCATCTCCACAGAAAGCCCTCATTTCTCTCTTCTGCTATCTTCTCCA 1495  
QY 1321 GACAGTGAACAAGGAGCTGCTGTTGAGGCTCTGACACACACACATCATCAGTCTGAT 1555  
DB 1496 GATAGTGAACAAGGAGCTGCTGTTGAGGCTCTGACACACACACATCATCAGTCTGAT 1555  
QY 1381 ACAATATGAAAAGCTGAGCCCCAAACAGGAGAGTGAACAGTGTAAAGTTTGAC 1440  
DB 1556 ACAATATGAAAAGCTGAGCCCCAAACAGGAGAGTGAACAGTGTAAAGTTTGAC 1615  
QY 1441 TACAGCAGAGAGAGAACTACAGCAACAATAAAGTGAACCTTTACAGATGAATGGAGC 1500  
DB 1616 TACAGCAGAGAGAGAACTACAGCAACAATAAAGTGAACCTTTACAGATGAATGGAGC 1675  
QY 1501 AGTATAGCATCAAGTTCTAGAGGAATTTGGAGGCTTGAACATCTGAGGGTCAGAGGAA 1560  
DB 1676 AGTATAGCATCAAGTTCTAGAGGAATTTGGAGGCTTGAACATCTGAGGGTCAGAGGAA 1735  
QY 1561 TCTTTCTGCTCCACAGAGCTCAGTGCACCAACAGAGAGAGAGTGAACAAAGCTCT 1620  
DB 1736 TCTTTCTGCTCCACAGAGCTCAGTGCACCAACAGAGAGAGAGTGAACAAAGCTCT 1795  
QY 1621 GAAGATTCATCAGAGGATGTGACAAATATATCAGGAAGGAGTATCTGCAGAAAACCCAGTT 1680

Db 1796 GAGAAATCATCAGAGGATGTCAGAAATATCAGGAAGGAGTATCTGCAGAAACCCAGTT 1855  
 QY 1681 GAGAACATATCAATTAACACCAATCAGATAAGTTCACACCCCAAGCCATTGGATTCCAAC 1740  
 Db 1856 GAGAACATATCAATTAACACCAATCAGATAAGTTCACACCCCAAGCCATTGGATTCCAAC 1915  
 QY 1741 TCAGGAGAAAGAAATCAGCTCAATCTTGATCGCTCTGTGGGTTCAGAGAAATCTGCT 1800  
 Db 1916 TCAGGAGAAAGAAATCAGCTCAATCTTGATCGCTCTGTGGGTTCAGAGAAATCTGCT 1975  
 QY 1801 TCATCTGAAAAAGCCCAAGGAACAGAAACTTCAGATCAGACTAGCAGTGTGCTTAC 1860  
 Db 1976 TCATCTGAAAAAGCCCAAGGAACAGAAACTTCAGATCAGACTAGCAGTGTGCTTAC 2035  
 QY 1861 AATGAAATTAACCAATCTCTGAGGCTCAGTTCCAAACAGAGCCACTGGGCTTCAGCT 1920  
 Db 2036 AATGAAATTAACCAATCTCTGAGGCTCAGTTCCAAACAGAGCCACTGGGCTTCAGCT 2095  
 QY 1921 CATGAAAGAAACATCCACAGGAGCTCTCTCTTCAGGACACAGATGACAGTGTGATGAC 1980  
 Db 2096 CATGAAAGAAACATCCACAGGAGCTCTCTCTTCAGGACACAGATGACAGTGTGATGAC 2155  
 QY 1981 CCAGTCTGATCCAGTGTGAGGTATCGAGCAGGACCTGGTGTATGACGCTCTGCTGTT 2040  
 Db 2156 CCAGTCTGATCCAGTGTGAGGTATCGAGCAGGACCTGGTGTATGACGCTCTGCTGTT 2115  
 QY 2041 GCCCGTATTCAGGAGTCTTCAGACGCGAGAAAGAAAGAAATGGAAGAAATTTGGAT 2100  
 Db 2216 GCCCGTATTCAGGAGTCTTCAGACGCGAGAAAGAAAGAAATGGAAGAAATTTGGAT 2275  
 QY 2101 ACTTTGAACATTAAGAGGCGCTAGTAAATGTTTATAAGGCCATCCACTCCAGG 2160  
 Db 2276 ACTTTGAACATTAAGAGGCGCTAGTAAATGTTTATAAGGCCATCCACTCCAGG 2335  
 QY 2161 ACAATGATAAAGAGCCCAATTTCTGGGTGCTAACTTTGTAATGAGTGTCTGACTGT 2220  
 Db 2336 ACAATGATAAAGAGCCCAATTTCTGGGTGCTAACTTTGTAATGAGTGTCTGACTGT 2295  
 QY 2221 GGCACATTTTCACTGGGATCGGCACATCTGCTGAGCATTGATGCTTCTGGAAGCTGAT 2280  
 Db 2396 GGCACATTTTCACTGGGATCGGCACATCTGCTGAGCATTGATGCTTCTGGAAGCTGAT 2455  
 QY 2281 AATCATGTGTTAACTGCTCGAGCCACATCTGCTGAGCATTGATGCTTCTGGAAGCTGAT 2340  
 Db 2456 AATCATGTGTTAACTGCTCGAGCCACATCTGCTGAGCATTGATGCTTCTGGAAGCTGAT 2515  
 QY 2341 ATAGATTATGACATAAAGATCTGTCACCATTAAGAGTCAAGGATTTTAAACCGAAA 2400  
 Db 2516 ATAGATTATGACATAAAGATCTGTCACCATTAAGAGTCAAGGATTTTAAACCGAAA 2575  
 QY 2401 CTGCTGATGAAGTTATTAACCTCGAAACCACTCATGCTGGAGAAACATGAGAAACCACTT 2460  
 Db 2576 CTGCTGATGAAGTTATTAACCTCGAAACCACTCATGCTGGAGAAACATGAGAAACCACTT 2635  
 QY 2461 ACAGTTCAGGCTCTTCACTGTTGAGGATGTTGCTTCACTTAATCATATCCGAGCTGAC 2520  
 Db 2636 ACAGTTCAGGCTCTTCACTGTTGAGGATGTTGCTTCACTTAATCATATCCGAGCTGAC 2695  
 QY 2521 CGGTGGAGGTTGACAGATCAGAGGCTCTGCTCAAGAGATGAAATGAGGATGAGGAA 2580  
 Db 2696 CGGTGGAGGTTGACAGATCAGAGGCTCTGCTCAAGAGATGAAATGAGGATGAGGAA 2755

## RESULT 3

AX078303  
 LOCUS AX078303 3022 bp DNA linear PAT 22-FEB-2001  
 DEFINITION Sequence 107 from Patent WO0107471.  
 ACCESSION AX078303  
 VERSION AX078303.1 GI:13157994  
 KEYWORDS human.  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 FEATURES  
 SOURCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 1 (bases 1 to 3022)  
 Hillman, J.L., Lal, P., Tang, Y.T., Yue, H., Au-Young, J., Bandman, O.,  
 Azimzai, Y., Yang, J., Lu, D.A., Baughn, M.R., Patterson, C. and Shah, P.  
 Cell cycle and proliferation, proteins  
 Patent: WO 0107471-A 107 01-FEB-2001;  
 Incyte Genomics, Inc. (US)

Location/Qualifiers  
 1..3022

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /note="Incyte ID No: 5682976CB1"

BASE COUNT 937 a 611 c 674 g 800 t  
 ORIGIN

## Query Match

Best Local Similarity 97.2%; Score 2508.4; DB 6; Length 3022;  
 Matches 2579; Conservative 0; Mismatches 1; Indels 60; Gaps 1;  
 QY 1 ATGTCCTCGGGTGGCTCTTACCCACACCTGTTGTGGGACGTGAGGAAAGTCCCTCGGG 60  
 Db 119 ATGTCCTCGGGTGGCTCTTACCCACACCTGTTGTGGGACGTGAGGAAAGTCCCTCGGG 178  
 QY 61 CTGGAGGACCCGTCCTCGGCTGCGAGTCTGCTACCTGCGGAGAGAGAAATTTATCCAAAGA 120  
 Db 179 CTGGAGGACCCGTCCTCGGCTGCGAGTCTGCTACCTGCGGAGAGAGAAATTTATCCAAAGA 238  
 QY 121 TTAACACTTGAAGCAACCCCTTAATGTGCATGATGGTGTGTTAATACAACTCTGTGGAAT 180  
 Db 239 TTAACACTTGAAGCAACCCCTTAATGTGCATGATGGTGTGTTAATACAACTCTGTGGAAT 298  
 QY 181 GACACTCGAGAAATATTTATCTGCTCAGATGACACCAAAATAGTAATAGTAATCTCT 240  
 Db 299 GACACTCGAGAAATATTTATCTGCTCAGATGACACCAAAATAGTAATAGTAATCTCT 358  
 QY 241 TACAGCAGAAAGGTTTGAACAACATTCGTTCAAGGACCCAGCAACATATTTAGTGCA 300  
 Db 359 TACAGCAGAAAGGTTTGAACAACATTCGTTCAAGGACCCAGCAACATATTTAGTGCA 418  
 QY 301 AACTCTTACCTTGTACAAATGATAACAGATTGTATCTCTCTCTGGAGATGAGTAATA 360  
 Db 419 AAGTTCTTACCTTGTACAAATGATAACAGATTGTATCTCTCTCTGGAGATGAGTAATA 478  
 QY 361 TTTTATACCAAGCTTGAGCAAGATGAGCAACCAACAGACATGCAATGCAATGCAATGCAAT 420  
 Db 479 TTTTATACCAAGCTTGAGCAAGATGAGCAACCAACAGACATGCAATGCAATGCAATGCAAT 538  
 QY 421 TATGNACTACTTATGAGATTATGACTGTACCAATGACCTTACACCTTACACCTTCTCTCTTGT 480  
 Db 539 TATGNACTACTTATGAGATTATGACTGTACCAATGACCTTACACCTTCTCTCTTGT 598  
 QY 481 GGTGAAGATGGAACCTGTTAGGTGTTTGTATACACCATCAAACTAGCTGCAACAAAGAA 540  
 Db 599 GGTGAAGATGGAACCTGTTAGGTGTTTGTATACACCATCAAACTAGCTGCAACAAAGAA 658  
 QY 541 GATTGTAAGATGATATTTTAACTCTGCGCTGCTGCACTGCTGCTGCTGCTGCTGCTGCTGCT 600  
 Db 659 GATTGTAAGATGATATTTTAACTCTGCGCTGCTGCACTGCTGCTGCTGCTGCTGCTGCTGCT 718  
 QY 601 CCACCAATACCATATTTACCTTCTCCCATCTTAATAAAGTCTGCGAGAGTACATCTCTGCTGT 780  
 Db 719 CCACCAATACCATATTTACCTTCTCCCATCTTAATAAAGTCTGCGAGAGTACATCTCTGCTGT 838  
 QY 721 GTTCCCGCTTTTATCTCTCCCATCTTAATAAAGTCTGCGAGAGTACATCTCTGCTGT 840  
 Db 839 GTTCCCGCTTTTATCTCTCCCATCTTAATAAAGTCTGCGAGAGTACATCTCTGCTGT 898  
 QY 781 TACAGTGAAGATGGTCAAGAGATTTCTGTTAGTTACTCTTTCAGATTACATATATCTTTT 840

[illegible]

```

/codon_start=1
/protein_id="CAC32385.1"
/db_xref="GI:13121649"
/translation="NSRGSYPHLLWVRKFLGLEDPRLRSYRLGRREFIORKLE
LPVNDKGVNTICWNDIGETILSGSDTKLIVSNPYSRKVLTTIRSHRANISAKF
GEDGVNFDRIKTSCKTKEDKDDILNCRRAATVAICPPPIYLAAGCSDSVRI
YDRMLGTRFRTAGTGTGVARFIPSHLNKSCRVTSLCYSEDGELLVSYSDY
IYLPDKDRTARELKTSAERREBELPOPPVKRLRLGCDMSDTPRPRPSERERDGE
QSPNVLMORMDMLSRWFEEASVAVOSNRGRSRERGGTSGSDISITLTPVPSDDL
EVSTAMEVTPAQFLOPQSTSSMSQAQUSTSPSPHSTPLLSPSDEQKQSVBA
SGHITHQSEFLRGPETALLKRLQQLKAAEQORQOELAAHTQQPSTSDQSHHG
SSQPHFADSPSSVNVKGLSMGLDEQNNNEKLSKPKGTPVLSLHYSTRETTTS
TKLVNDEWSSIASSNGSGSESEFVPOSSVOPPEGDESKAPERSSEVY
KEYTQGSYAENPVNHNITQSDKFTAKPLDSNSGERNDLNLDSRGVPEASESSEKA
IPKARYSQTSEATNNNTNPPFQTEATGPSAHEETSTRDSALQDDSDSDDDPVL
MIKANFWPAGFVMTGSEGHIFIDWHTAHLMLLEADNHVVCILQHPFPILASS
GIDYDKINSPLESIRFNKRLADEVITRNELMLEETRNITVPSFMRMLASLNIH
RADRLGDRSGSGQENENEDEP"
BASE COUNT 1036 a 723 c 778 g 883 t
ORIGIN

```

```

Query Match          90.4%; Score 2332.6; DB 6; Length 3420;
Best Local Similarity 91.6%; Pred. No. 0;
Matches 2576; Conservative 0; Mismatches 4; Indels 231; Gaps 1;

QY 1 ATGTCGGGGTGGCCCTACCCACACCTGTGTCGGACGTGAGGAAAAGTCCCTCGG 60
DB 176 ATGTCGGGGTGGCCCTACCCACACCTGTGTCGGACGTGAGGAAAAGTCCCTCGG 60
QY 61 CTGGAGACCCCTCGCGGTGGGAGTGCCTACCTGGGAAGAAGAGATTTATCCAAAGA 120
DB 236 CTGGAGACCCCTCGCGGTGGGAGTGCCTACCTGGGAAGAAGAGATTTATCCAAAGA 120
QY 121 TTAACACTTGAAGCAACCCCTTAATGTGATGATGCTTGTGTAATACAACTGTGGAA 180
DB 296 TTAACACTTGAAGCAACCCCTTAATGTGATGATGCTTGTGTAATACAACTGTGGAA 180
QY 181 GACACTGGAGAAATATTTATCTGGCTGACATGACACCAATAGTAATAGTAATCCCT 240
DB 356 GACACTGGAGAAATATTTATCTGGCTGACATGACACCAATAGTAATAGTAATCCCT 240
QY 241 TACAGCAGAAAGTTTGACAAACATTCGTTACGGGACCCAGCAACACATATTAGTGCA 300
DB 416 TACAGCAGAAAGTTTGACAAACATTCGTTACGGGACCCAGCAACACATATTAGTGCA 300
QY 301 AAGTCTTACCTGTACAAATGATAACAGATTGTATCTGCTGAGATGGAGTAATA 360
DB 476 AAGTCTTACCTGTACAAATGATAACAGATTGTATCTGCTGAGATGGAGTAATA 360
QY 361 TTTTATACCAACGTTGAGCAAGATGCAAGAACCAACAGACAAATGATGCTGCTGAGATGAGTAATA 535
DB 536 TTTTATACCAACGTTGAGCAAGATGCAAGAACCAACAGACAAATGATGCTGCTGAGATGAGTAATA 535
QY 421 TATGGAACCTACTTATCAGATATTGACTGTACCAATGACCCCTTACACATTTCTCTCTGT 480
DB 596 TATGGAACCTACTTATCAGATATTGACTGTACCAATGACCCCTTACACATTTCTCTCTGT 480
QY 481 GGTGAGATGGAACTGTTAGTGGTTTGATACACGCATCAAACTAGCTGCACAAAAGAA 540
DB 656 GGTGAGATGGAACTGTTAGTGGTTTGATACACGCATCAAACTAGCTGCACAAAAGAA 540
QY 541 GATTGTAAGATGATATTTTAAATTAACCTGTCAGCTGCTGCCAGCTGTGTTGCTATTGTC 715
DB 716 GATTGTAAGATGATATTTTAAATTAACCTGTCAGCTGCTGCCAGCTGTGTTGCTATTGTC 715
QY 601 CCACCAATACCAATATTACCTTGTGTTGGTTTGTGACAGCTCAGTACGAAATATGAT 660
DB 776 CCACCAATACCAATATTACCTTGTGTTGGTTTGTGACAGCTCAGTACGAAATATGAT 660
QY 661 CGGGAATGCTGGCACAAGAGCTACAGGGAATTTATGAGGTCGAGGAGCTACTTGGAA 720
DB 720 CGGGAATGCTGGCACAAGAGCTACAGGGAATTTATGAGGTCGAGGAGCTACTTGGAA 720

```

```

DB 836 CGGCGAATGCTGGGCACAAGAGCTACAGGGAATTAATGAGGTCGAGGACTACTGGAATG 895
QY 721 GTTGCCCGTTTATTTCCCTTCCCATCTTAATAATAGTCTCTCAGAGTGCACATCTCTGTGT 895
DB 896 GTTGCCCGTTTATTTCCCTTCCCATCTTAATAATAGTCTCTCAGAGTGCACATCTCTGTGT 895
QY 781 TACAGTGAAGATGCTCAAGAGATTCTCGTGTAGTACTCTTCCAGATTACATATATCTTTT 840
DB 956 TACAGTGAAGATGCTCAAGAGATTCTCGTGTAGTACTCTTCCAGATTACATATATCTTTT 840
QY 841 GACCCGAAAGATGATACAGCACCAGAACTTAAACCTCTCTCGGGAAGAGAGAGAA 900
DB 1016 GACCCGAAAGATGATACAGCACCAGAACTTAAACCTCTCTCGGGAAGAGAGAGAA 1075
QY 901 GAGTTCGGCACAACCCACAGTTAAAGCGTTTGAAGACTTTCGTTGGTATTTGGTTCAGATCTGA 960
DB 1076 GAGTTCGGCACAACCCACAGTTAAAGCGTTTGAAGACTTTCGTTGGTATTTGGTTCAGATCTGA 1135
QY 961 CCCAGAGCAAGCGCGGAGAGTGAACGAGAGAGATGGAGAGCAGAGTCCCAATGTGCA 1020
DB 1136 CCCAGAGCAAGCGCGGAGAGTGAACGAGAGAGATGGAGAGCAGAGTCCCAATGTGCA 1195
QY 1021 TTGATCCAGAGATGCTGTGATATGTTATCAAGATGTTTGAAGAGCAAGTGAAGTTCGA 1080
DB 1196 TTGATCCAGAGATGCTGTGATATGTTATCAAGATGTTTGAAGAGCAAGTGAAGTTCGA 1255
QY 1081 CAAAGCAATAGAGGACGAGGAGATCTGACCCAGAGAGTGGAGAGTGAACAAAGTCAATGATTT 1140
DB 1256 CAAAGCAATAGAGGACGAGGAGATCTGACCCAGAGAGTGGAGAGTGAACAAAGTCAATGATTT 1315
QY 1141 TCAACTCTTCTAGGTCCTCCATCAAGTCTGATTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1200
DB 1316 TCAACTCTTCTAGGTCCTCCATCAAGTCTGATTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1375
QY 1201 GTAGATCTCCAGCTGAACAAATTTCTTCAAGCTCTTACATCTTACATCTTACATCTTACATCT 1260
DB 1376 GTAGATCTCCAGCTGAACAAATTTCTTCAAGCTCTTACATCTTACATCTTACATCTTACATCT 1335
QY 1261 GCTCATTTCGACATCTCTCCACAGAAAGCCCTCAATCTACTCTTCTTCTTCTTCTTCTTCTTCT 1377
DB 1436 GCTCATTTCGACATCTCTCCACAGAAAGCCCTCAATCTACTCTTCTTCTTCTTCTTCTTCTTCT 1495
QY 1321 GACAGTGAACAAAGGAGTCTGTTGAGGATCTGACACACACACACATCATCAGTCT --- 1377
DB 1496 GATAGTGAACAAAGGAGTCTGTTGAGGATCTGTTGAGGATCTGTTGAGGATCTGTTGAGGATCT 1555
QY 1378 --- 1377
DB 1556 TTTTAAAGGGGCGCTGAGATAGCTTTGCTTCTGTAAGCGCTGCAACAACTGAGGCTTAAG 1615
QY 1378 --- 1377
DB 1616 AAGGCTGAGCAGCAGAGGAGCAGCAGAGCTAGCTGCACATACCCAGCAACAGGATTAAG 1675
QY 1378 --- 1377
DB 1676 TCTGATCAGTCTTCTCAGAGGGCTCTTACAGGACCTCATGCTTCTCAGATTCCTCTTCT 1735
QY 1378 --- 1377
DB 1736 TCTGTGTTAAACAAAGCTCGGATCCATGCTCAGCTTACAGCAGCAACAGGATTAACAAAT 1795
QY 1390 GAAAGCTGAGCCCAACCCAGGACAGGAGTGAACACAGTTTAAAGTTTGAAGTGAAGTGAAGTGAAG 1449
DB 1796 GAAAGCTGAGCCCAACCCAGGACAGGAGTGAACACAGTTTAAAGTTTGAAGTGAAGTGAAGTGAAG 1855
QY 1450 GAGGAACTACTACAGCAATATAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1509
DB 1856 GAGGAACTACTACAGCAATATAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1915
QY 1510 TCAAGTCTTAGAGGATTTGGGAGGATTCGAAATCTGAGGTCGAGGAGTGAAGTGAAGTGAAGTGAAG 1569
DB 1916 TCAAGTCTTAGAGGATTTGGGAGGATTCGAAATCTGAGGTCGAGGAGTGAAGTGAAGTGAAGTGAAG 1975

```







QY	1378	-----	1377
Db	1676	TCTGATCAGTCTTCTCATGAGGGCTTTCACAGGACCTCATGCTTCAGATCTCTCTTCT	1735
QY	1378	-----	
Db	1736	TCTGTGGTTAACAAACAGCTCGGATCCATGTCACGAGCAGCAACAGGATACAAATAAT	1389
QY	1390	GAAGAAGCTGAGCCCCAAACACAGGACAGGTGAACCAAGTTTTAAAGTTTGACACTACAGCACA	1795
Db	1796	GAAGAAGCTGAGCCCCAAACACAGGACAGGTGAACCAAGTTTTAAAGTTTGACACTACAGCACA	1449
QY	1450	GAAGGAACAACTACAAAGCACAATAAAACCTGAACTTTACAGATGAATGGAGCAGTATAGCA	1855
Db	1856	GAAGGAACAACTACAAAGCACAATAAAACCTGAACTTTACAGATGAATGGAGCAGTATAGCA	1509
QY	1510	TCAAAGTCTAGAGGAATGGGAGCCATTCGCAATCTGAGGGTCAGGAGGAATCTTTCTGTC	1915
Db	1916	TCAAAGTCTAGAGGAATGGGAGCCATTCGCAATCTGAGGGTCAGGAGGAATCTTTCTGTC	1569
QY	1570	CCACAGAGCTCAGTCACACACAGAGGAGACAGTGAACAAAGCTCTCTGAAGATCA	1629
Db	1976	CCACAGAGCTCAGTCACACACAGAGGAGACAGTGAACAAAGCTCTCTGAAGATCA	2035
QY	1630	TCAGAGGATGTGACAAATATACGAAAGGAGTATCTCCAGAAACCCAGTTGAGAACCAT	1689
Db	2036	TCAGAGGATGTGACAAATATACGAAAGGAGTATCTCCAGAAACCCAGTTGAGAACCAT	2095
QY	1690	ATCAATATAACACAACTCAGATAGTTTCACAGCCAGCCATTGGATTCACCACTCAGAGAA	1749
Db	2096	ATCAATATAACACAACTCAGATAGTTTCACAGCCAGCCATTGGATTCACCACTCAGAGAA	2155
QY	1750	AGAAATGACCTCAATCTTGATCGCTCTGTGGGTTCACAGAAATCTGCTTCATCTGAA	1809
Db	2156	AGAAATGACCTCAATCTTGATCGCTCTGTGGGTTCACAGAAATCTGCTTCATCTGAA	2215
QY	1810	AAAGCCAAAGAAACACAGAACTTCAGATCAGACTAGCAGTGCACAGAGTGCACCAATGAAAT	1869
Db	2216	AAAGCCAAAGAAACACAGAACTTCAGATCAGACTAGCAGTGCACAGAGTGCACCAATGAAAT	2275
QY	1870	ACACCAATCTTGAGCCCTCAGTTCACAAACAGAGCCAGTGGCCCTTCAGCTCATGAAGAA	1929
Db	2276	ACACCAATCTTGAGCCCTCAGTTCACAAACAGAGCCAGTGGCCCTTCAGCTCATGAAGAA	2335
QY	1930	ACATCCACAGGGACTCTGCTCTTCAGGACACAGATGACAGTGCATGATGATGACCCAGTCTG	1989
Db	2336	ACATCCACAGGGACTCTGCTCTTCAGGACACAGATGACAGTGCATGATGATGACCCAGTCTG	2395
QY	1990	ATCCAGGTGCAAGGTATCGAGCAGGACCTGGTGATAGCCCTCTGCTGTGGCCGATAT	2049
Db	2396	ATCCAGGTGCAAGGTATCGAGCAGGACCTGGTGATAGCCCTCTGCTGTGGCCGATAT	2455
QY	2050	CAGGAGTCTTCAGACGGAGAAAGAAAGAAATGGAAGAAATGGATATCTTTGAAC	2109
Db	2456	CAGGAGTCTTCAGACGGAGAAAGAAAGAAATGGAAGAAATGGATATCTTTGAAC	2515
QY	2110	ATTAGAAGCCCTAGTAAATGGTTTAAAGGCCATCGCAACTCCAGGACAAATGATA	2169
Db	2516	ATTAGAAGCCCTAGTAAATGGTTTAAAGGCCATCGCAACTCCAGGACAAATGATA	2575
QY	2170	AAAGAAGCCAAPTTCTGGGGTGCTAACTTTCTGATGAGTGGTTCGACTGTGGCCACATT	2229
Db	2576	AAAGAAGCCAAPTTCTGGGGTGCTAACTTTCTGATGAGTGGTTCGACTGTGGCCACATT	2635
QY	2230	TTCACTCTGGGATCGGCACACTGCTGAGCATTTGATGCTCTGGAAGCTGATATCATGTG	2289
Db	2636	TTCACTCTGGGATCGGCACACTGCTGAGCATTTGATGCTCTGGAAGCTGATATCATGTG	2695
QY	2290	GTAACCTGCCTGCAGCCACATCCGTTTGACCCAAATTTAGCCTCATCTGGCATAGATTAT	2349
Db	2696	GTAACCTGCCTGCAGCCACATCCGTTTGACCCAAATTTAGCCTCATCTGGCATAGATTAT	2755
QY	2350	GACATAAAGATCTGGTCCACATTTAGAAGAGTCAAGGATTTTTAAACCGAAACCTGCTGAT	2409

[illegible][illegible]



QY	1907	CTGGGCTTCAGCTCATGAAGAAACATCACCAGGAGTCTGCTCTTTCAGGACACAGATG	1966
Db	1755	CTGGGCTTCAGCTCATGAAGAAACATCACCAGGAGTCTGCTCTTTCAGGACACAGATG	1813
QY	1967	ACAGTGATGATGACCCAGTCTGATCCAGGTGACAGGTATCGAGCAGACCTGGTGATA	2026
Db	1814	ACAGTGATGATGACCCAGTCTGATCCAGGTGACAGGTATCGAGCAGACCTGGTGATA	1870
QY	2027	GAGCTCTGCTGTGCTGCTGATTCAGGAGTCTTCAGAGCGGAGAAAGAAAGAAAGAA	2086
Db	1871	GAGCTCTGCTGTGCTGCTGATTCAGGAGTCTTCAGAGCGGAGAAAGAAAGAAAGAA	1930
QY	2087	TGGAAGAAATGGTACTTTGAACATTAAGAGCGGCTAGTAAATAAGTTTATAAAGGCC	2146
Db	1931	TGGAAGAAATGGTACTTTGAACATTAAGAGCGGCTAGTAAATAAGTTTATAAAGGCC	1990
QY	2147	ATCGCAACTCCAGGACATGATTAAGAGCGGCTAGTAAATAAGTTTATAAAGGCC	2206
Db	1991	ATCGCAACTCCAGGACATGATTAAGAGCGGCTAGTAAATAAGTTTATAAAGGCC	2050
QY	2207	GTGGTCTGACTGTGGCCACATTTTCATCTGGGATCGGCACACTGCTGAGCATTTGATGC	2266
Db	2051	GTGGTCTGACTGTGGCCACATTTTCATCTGGGATCGGCACACTGCTGAGCATTTGATGC	2110
QY	2267	TTCTGGAAGCTGATTAATCATGTGGTAAACTGCTGAGCCACATCCGTTGACCCAAATTT	2326
Db	2111	TTCTGGAAGCTGATTAATCATGTGGTAAACTGCTGAGCCACATCCGTTGACCCAAATTT	2170
QY	2327	TAGCCCTCATCTGGCAGATGATTAAGAGCGGCTAGTAAATAAGTTTATAAAGGCC	2386
Db	2171	TAGCCCTCATCTGGCAGATGATTAAGAGCGGCTAGTAAATAAGTTTATAAAGGCC	2230
QY	2387	TTTCTAACCGAAACCTGCTGATGAAGTTTATAAAGCGGCTAGTAAATAAGTTTATAAAGGCC	2446
Db	2231	TTTCTAACCGAAACCTGCTGATGAAGTTTATAAAGCGGCTAGTAAATAAGTTTATAAAGGCC	2290
QY	2447	CTGAAACACCATTTACAGTTCAGGCTCTTTCATGTTGAGGATGTTGGCTTCACTTAATC	2506
Db	2291	CTGAAACACCATTTACAGTTCAGGCTCTTTCATGTTGAGGATGTTGGCTTCACTTAATC	2350
QY	2507	ATATCCGAGCTGACCGGTGGAGGCTGACAGATCAGAAAGCTCTGTCAGAGAAATGAAA	2566
Db	2351	ATATCCGAGCTGACCGGTGGAGGCTGACAGATCAGAAAGCTCTGTCAGAGAAATGAAA	2410
QY	2567	ATGAGGATGAGGAA	2580
Db	2411	ATGAGGATGAGGAA	2424
RESULT 8			
AK074378/c			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
QY	884	CGAAGAGAGAGAGAGAGTTCGACAAACACACAGTTAAAGCGTTTGAGACTTCGTGGTG	943
Db	675	CGAAGAGAGAGAGAGAGTTCGACAAACACACAGTTAAAGCGTTTGAGACTTCGTGGTG	734
QY	944	ATTGCTCAGATCTGACCCAGAGCAAGCGCGGAGAGTGAACGAGAGAGAGAGAGAGC	1003
Db	735	ATTGCTCAGATCTGACCCAGAGCAAGCGCGGAGAGTGAACGAGAGAGAGAGAGAGC	794
QY	1004	AGAGTCCCAATGTTCATGTAGTCAGAGAAATGCTGATATGTTTATCAAGATGGTTGAAG	1063
Db	795	AGAGTCCCAATGTTCATGTAGTCAGAGAAATGCTGATATGTTTATCAAGATGGTTGAAG	854
QY	1064	AAGCAAGTGAAGTTCGACAAAGCAATAGAGAGAGAGAGAGTCTCGACCCAGAGGTGAA	1123
Db	855	AAGCAAGTGAAGTTCGACAAAGCAATAGAGAGAGAGAGTCTCGACCCAGAGGTGAA	914
QY	1124	CAAGTCATCAGATATTTCAACTCTTCTAGGCTCCATCAAGTCTGATTTGGAAGTGA	1183
Db	915	CAAGTCATCAGATATTTCAACTCTTCTAGGCTCCATCAAGTCTGATTTGGAAGTGA	974
QY	1184	GTGAACTGCAATGGAGTACTGACCTGAAACAAATTTCTTACGCTTCTTACATCTCT	1243
Db	975	GTGAACTGCAATGGAGTACTGACCTGAAACAAATTTCTTACGCTTCTTACATCTCT	1034
QY	1244	CTACATGTCAGCTCAGGCTCATTGAGATCATCTCCACAGAAAGCCCTCATCTTACTC	1303
Db	1035	CTACATGTCAGCTCAGGCTCATTGAGATCATCTCCACAGAAAGCCCTCATCTTACTC	1094
QY	1304	CTTTGCTATCTTCTCCAGAGTGAACAAAGGAGTCTGTTGAGGCTATCTGGACACACA	1363
Db	1095	CTTTGCTATCTTCTCCAGAGTGAACAAAGGAGTCTGTTGAGGCTATCTGGACACACA	1154
QY	1364	CACATCATCAGTCTGATTAACAAATATGAAAGCTGAGC	1401
Db	1155	CACATCATCAGTCTGATTAACAAATATGAAAGCTGAGC	1214
QY	1402	-----CCCAACACAGGAGTGAACCAAGCTGAGCCAGGAGTGAACCAAG	1426
Db	1215	TTGACGAGCAACAGGATACCAATGAAGAAAGCTGAGCCAGGAGTGAACCAAG	1274
QY	1427	TTTAAAGTTTGACTACAGACAGAGAGGAAACATACAGAGCAATTAACCTGAACTTTA	1486
Db	1275	TTTAAAGTTTGACTACAGACAGAGAGGAAACATACAGAGCAATTAACCTGAACTTTA	1334
QY	1487	CAGATGAATGGAGCAGTATGACATCAAGTTCTAGAGAAATTTGGAGCCCATTCGAATCTG	1546
Db	1335	CAGATGAATGGAGCAGTATGACATCAAGTTCTAGAGAAATTTGGAGCCCATTCGAATCTG	1394
QY	1547	AGGCTCAGGAGGAAATCTTCTGCTCCACAGAGCTCAGTGAACCAACAGGAGAGAGAGT	1606
Db	1395	AGGCTCAGGAGGAAATCTTCTGCTCCACAGAGCTCAGTGAACCAACAGGAGAGAGAGT	1454
QY	1607	AAACAAAGCTCTTGAGAAATCATCAGAGGATGACAAATAATATCAGGAGGAGTATCTG	1666
Db	1455	AAACAAAGCTCTTGAGAAATCATCAGAGGATGACAAATAATATCAGGAGGAGTATCTG	1514
QY	1667	CAGAAACCCAGTTGAGAAACCATATCAATATAACAAATCAGATGATGATGATGATGATG	1726
Db	1515	CAGAAACCCAGTTGAGAAACCATATCAATATAACAAATCAGATGATGATGATGATGATG	1574
QY	1727	CATTGGATTCACCTCAGGAGAGAAAGATGACCTCAATCTTGTGATGCTCTTGTGGGGTTC	1786
Db	1575	CATTGGATTCACCTCAGGAGAGAAAGATGACCTCAATCTTGTGATGCTCTTGTGGGGTTC	1634
QY	1787	CAGAAGAAATCTGCTCATCTGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGC	1846
Db	1635	CAGAAGAAATCTGCTCATCTGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGC	1694
QY	1847	CTGAGAGTGTACCAATGAATAACCAACCTCCTGAGCCTCAGTTTCCAAACAGAGGCA	1906
Db	1695	CTGAGAGTGTACCAATGAATAACCAACCTCCTGAGCCTCAGTTTCCAAACAGAGGCA	1754

Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan  
(E-mail: [cdnal@ims.u-tokyo.ac.jp](mailto:cdnal@ims.u-tokyo.ac.jp), Tel: 81-3-5449-5286,  
Fax: 81-3-5449-5416)

**COMMENT**

2000 human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction: 5'- and 3'-end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency). Location/Qualification

**FEATURES**  
**source**

```

BASE COUNT
ORIGIN
          546 a      549 c      530 g      700 t
          /organism="Homo sapiens"
          /db_xref="taxon:9606"
          /clone="HEP22180"
          /cell_line="HepG2"
          /cell_type="hepatoma"
          /clone.lib="HEP"
          /note="Cloning vector: pME18SF13"

```

Query Match 80.5%; Score 2076.4; DB 9; Length 2325;  
 Best Local Similarity 99.7%; Pred. No. 0;  
 Matches 2080; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	52
--	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----

QY	721	GTGGCCCGTTTTATTTCTCTCCCATCTTAATAATAAAGTCCTGCAGATGACATCTCTGTGT	780
DB	1370	GTGGCCCGTTTTATTTCTCTCCCATCTTAATAATAAAGTCCTGCAGATGACATCTCTGTGT	1311
QY	781	TACAGTGAAGATGGTCAAGAGATCTCTGTTAGTTACTCTTCAGATTACATATATCTTTTT	840
DB	1310	TACAGTGAAGATGGTCAAGAGATCTCTGTTAGTTACTCTTCAGATTACATATATCTTTTT	1251
QY	841	GACCCGAAGATGATACAGCAGCAAGAACTTAAACTCCTTCTGCGGAAGAGAGAAGAA	900
DB	1250	GACCCGAAGATGATACAGCAGCAAGAACTTAAACTCCTTCTGCGGAAGAGAGAAGAA	1191
QY	901	GAGTTGGCACAACACACAGTTAAAGCGTTTGAGACTTCGTGGTGAATTTGTCAGATACTGA	960
DB	1190	GAGTTGGCACAACACACAGTTAAAGCGTTTGAGACTTCGTGGTGAATTTGTCAGATACTGA	1131
QY	961	CCCAGACAAGCCGGGAGAGTGAACGAGAAGAGATGAGAGCAGAGTCCCAATGTGCA	1020
DB	1130	CCCAGACAAGCCGGGAGAGTGAACGAGAAGAGATGAGAGCAGAGTCCCAATGTGCA	1071
QY	1021	TTGATGCAGAGAATGCTGTGATGTTATCAAGATGGTTTGAAGAACGAACTGAGTTGCA	1080
DB	1070	TTGATGCAGAGAATGCTGTGATGTTATCAAGATGGTTTGAAGAACGAACTGAGTTGCA	1011
QY	1081	CAAAGCAATAGAGCAGGAGAAGATCTCGACCCAGAGGTGGAACAAGTCAATCAGATATT	1140
DB	1010	CAAAGCAATAGAGCAGGAGAAGATCTCGACCCAGAGGTGGAACAAGTCAATCAGATATT	951
QY	1141	TCAACTCTTCTAGGTCGCCATCAAGTCTGTGATTTGGAAGTGAAGTGAAGTCAATCGAA	1200
DB	950	TCAACTCTTCTAGGTCGCCATCAAGTCTGTGATTTGGAAGTGAAGTGAAGTCAATCGAA	891
QY	1201	GTAGATACTCCAGCTGAACAATTTCTTCAGCCTTCTACATCTCTTCAATGTCAAGTCTCAG	1260
DB	890	GTAGATACTCCAGCTGAACAATTTCTTCAGCCTTCTACATCTCTTCAATGTCAAGTCTCAG	831
QY	1261	GCTCATTCGCATCATCTCCCAAGAAAGCCCTCATTTCTACTCTTCTTCTCTCTCTCTCA	1320
DB	830	GCTCATTCGCATCATCTCCCAAGAAAGCCCTCATTTCTACTCTTCTTCTCTCTCTCTCA	771
QY	1321	GACAGTGAACAAGAGCAGTCTGTTGAGCAGCATCTGCACACCACACACATCATCAGTCTGAT	1380
DB	770	GACAGTGAACAAGAGCAGTCTGTTGAGCAGCATCTGCACACCACACACATCATCAGTCTGAT	711
QY	1381	AACAATAATGAAGAGCTGAGCCCAACAGGAGCAGGTGAACCAAGTTTAAAGTTTGAC	1440
DB	710	AACAATAATGAAGAGCTGAGCCCAACAGGAGCAGGTGAACCAAGTTTAAAGTTTGAC	651
QY	1441	TACAGCACAGAAGAAACAATACAGACACAATAAAGTGAACCTTACAGATGAATGGAGC	1500
DB	650	TACAGCACAGAAGAAACAATACAGACACAATAAAGTGAACCTTACAGATGAATGGAGC	591
QY	1501	AGTATAGATCAAGTTCTAGAGGAATGGAGCCATTGCAATCTGAGGGTCAGAGGAA	1560
DB	590	AGTATAGATCAAGTTCTAGAGGAATGGAGCCATTGCAATCTGAGGGTCAGAGGAA	531
QY	1561	TCCTTTGCTCCACAGAGCTCAGTGCACACCAAGAGCAGTGAACAAAGCTCCT	1620
DB	530	TCCTTTGCTCCACAGAGCTCAGTGCACACCAAGAGCAGTGAACAAAGCTCCT	471
QY	1621	GAAGAATCATCAGAGATGTGACAAATATCAGGAAGGAGTATCTGCAGAAAACCCAGTT	1680
DB	470	GAAGAATCATCAGAGATGTGACAAATATCAGGAAGGAGTATCTGCAGAAAACCCAGTT	411
QY	1681	GAGAACCATATCAATATACACAATCAGATAGTTTACAGGCCAAGCCATTTGGATTCCAAC	1740
DB	410	GAGAACCATATCAATATACACAATCAGATAGTTTACAGGCCAAGCCATTTGGATTCCAAC	351
QY	1741	TCAGAGAAAGAAATGACCTCAATCTTGATCGCTCTTGTGGGGTTCACAGAAATCTGCT	1800
DB	350	TCAGAGAAAGAAATGACCTCAATCTTGATCGCTCTTGTGGGGTTCACAGAAATCTGCT	291
QY	1801	TCATCTCTGAAAAGCCAAAGGAACCAAGAACTTCAGATCAGCTAGCACTGAGAGTGTCTACC	1860





```
Db 1081 CGAGCAACAGGATAAATAAATGAAAGCTGAGCCCAACAGGACAGGTGAACAGT 1140
QY 1428 TTAAAGTTTGCACTACAGCACAGAGGACAACTACAGCAATAAAACTGAACCTTAC 1487
Db 1141 TTAAAGTTTGCACTACAGCACAGAGGACAACTACAGCAATAAAACTGAACCTTAC 1200
QY 1488 AGATGAATGAGGACGATATAGCATCAAGTTCTAGAGAAATTGGAGCAATTGCAATCTGA 1547
Db 1201 AGATGAATGAGGACGATATAGCATCAAGTTCTAGAGAAATTGGAGCAATTGCAATCTGA 1260
QY 1548 GGGTCAGGAGGAATCTTTCGTCACAGAGCTCACTGCAACCAACAGAGGACAGTGA 1607
Db 1261 GGGTCAGGAGGAATCTTTCGTCACAGAGCTCACTGCAACCAACAGAGGACAGTGA 1320
QY 1608 AACAAAGCTCCTGAAGATCATCAGAGGATGTGCAAAATATCAGGAGGATCTGCG 1667
Db 1321 AACAAAGCTCCTGAAGATCATCAGAGGATGTGCAAAATATCAGGAGGATCTGCG 1380
QY 1668 AGAAACCCAGTTGAGAACCATATCAATATTAACACATCAGATAAGTTTCAGAGCAAGCC 1727
Db 1381 AGAAACCCAGTTGAGAACCATATCAATATTAACACATCAGATAAGTTTCAGAGCAAGCC 1440
QY 1728 ATTGGATTCCAATCAGGAGAAAGAAATGACCTCAATCTTGATCGCTTGTGGGTTCC 1787
Db 1441 ATTGGATTCCAATCAGGAGAAAGAAATGACCTCAATCTTGATCGCTTGTGGGTTCC 1500
QY 1788 AGAAGATCTGCTCACTCTCAAAAGCCCAAGAACCAAGAACTTCAGATCAGATGACAC 1847
Db 1501 AGAAGATCTGCTCACTCTCAAAAGCCCAAGAACCAAGAACTTCAGATCAGATGACAC 1560
QY 1848 TGAGAGTGCTPACCAATGAAATTAACCAATCTGAGCCCTCACTGAGCCCTCACTGAGCC 1907
Db 1561 TGAGAGTGCTPACCAATGAAATTAACCAATCTGAGCCCTCACTGAGCCCTCACTGAGCC 1620
QY 1908 TGGGCTCTCAGCTCATGAGAAACATCCACAGGAGTCTGCTCTTCAGGACACAGATGA 1967
Db 1621 TGGGCTCTCAGCTCATGAGAAACATCCACAGGAGTCTGCTCTTCAGGACACAGATGA 1680
QY 1968 CAGTGATGATGACCCAGTCTGATCCAGGTGCAAGGTATCAGGACGAGGACCTGGTATAG 2027
Db 1681 CAGTGATGATGACCCAGTCTGATCCAGGTGCAAGGTATCAGGACGAGGACCTGGTATAG 1740
QY 2028 AGGCTCTGCTGTCGCTTATCAGAGTCTTTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2087
Db 1741 AGGCTCTGCTGTCGCTTATCAGAGTCTTTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1800
QY 2088 GGAAGATTTGATCTTTGAACATAGAGGCGCTAGTAAATGCTTTTAAAGGCCA 2147
Db 1801 GGAAGATTTGATCTTTGAACATAGAGGCGCTAGTAAATGCTTTTAAAGGCCA 1860
QY 2148 TCGCAACTCCAGGACATGATAAAGAGCCAAATTCCTGGGCTGCTACTTTGTAATGAG 2207
Db 1861 TCGCAACTCCAGGACATGATAAAGAGCCAAATTCCTGGGCTGCTACTTTGTAATGAG 1920
QY 2208 TGGTCTGACTGTGGCCACATTTTCATCTGGGATCGGACACTGCTGAGCAATTTGATGCT 2267
Db 1921 TGGTCTGACTGTGGCCACATTTTCATCTGGGATCGGACACTGCTGAGCAATTTGATGCT 1980
QY 2268 TCTGGAAGCTGATAATCATGTGTGTAATCTGCTGACCCACATCCGTTTGACCCCAATTTT 2327
Db 1981 TCTGGAAGCTGATAATCATGTGTGTAATCTGCTGACCCACATCCGTTTGACCCCAATTTT 2040
QY 2328 AGCCTCATCTGGCATAGATTATGACATAAAGATCTGGTCAACATTAAGAGATCAAGGAT 2387
Db 2041 AGCCTCATCTGGCATAGATTATGACATAAAGATCTGGTCAACATTAAGAGATCAAGGAT 2100
QY 2388 TTTTAAACCGAAACTTGTGATGAAGTTTAACTCGAAAGCAACTCATGCTGCGAAGAAC 2447
Db 2101 TTTTAAACCGAAACTTGTGATGAAGTTTAACTCGAAAGCAACTCATGCTGCGAAGAAC 2160
QY 2448 TAGAAACACATTTACAGTTCCAGCCTCTTTCATGTTGAGGATGTTGGCTCTCACTTAATCA 2507
```

```
Db 2161 TAGAAACACATTACAGTTCAGCCTCTTTCATGTTGAGGATGTTGGCTTCACTTAATCA 2220
QY 2508 TATCCAGCTGACCGGTTGGAGGTTGACAGATCAGAGGCTCTGGTCAAGAGATGAAAA 2567
Db 2221 TATCCAGCTGACCGGTTGGAGGTTGACAGATCAGAGGCTCTGGTCAAGAGATGAAAA 2280
QY 2568 TGAGGATGAGGAA 2580
Db 2281 TGAGGATGAGGAA 2293
```

```
RESULT 10
BC025262
LOCUS
DEFINITION
Homo sapiens, Similar to PC326 protein, clone MGC:39152
IMAGE:4763336, mRNA, complete cds.
ACCESSION
BC025262
VERSION
BC025262.1 GI:19263812
KEYWORDS
MGC.
SOURCE
Homo sapiens.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 2024)
AUTHORS
Strausberg, R.
TITLE
Direct Submission
JOURNAL
Submitted (05-MAR-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK
COMMENT
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabbs@mail.nih.gov
Tissue Procurement: Louis Staudt
cDNA Library Preparation: Rubin Staudt
DNA Sequencing Arrayed by: The I.M.A.G.E. Consortium (LLNL)
BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Gulin,
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
Schein, Duane Smalrus, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranda Tsai, Natasja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.
```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAL plate: 42 Row: 9 Column: 3  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 8923955.

## FEATURES

## source

location/Qualifiers

1..2024

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="MGC:39152 IMAGE:4763336"

/tissue\_type="Primary B-Cells from Tonsils"

/clone\_lib="NIH\_MGC\_48"

/lab\_host="DH10B-R"

/note="Vector: pOTB7"

58..1599

/codon\_start=1

/product="Similar to PC326 protein"

/protein\_id="AAH25262.1"

/db\_xref="GI:19263813"

/translation="MLSRWFEEASEVAQSNRGRSRPRGGTQSQDITLPTVPSSPD

ASGHHTHQSDNNNEKLSFGPGTGPVLSHYSTGGTSTIKLFTDWSIASRR

GTGSHCKSEGESEFVSQVPPGDSKAPESSEDATKIQGVSAENVEHNH

ITQSDKFTAKPLDSDNSGRNDLDRSCGVPESSAKAKPEPSTDSATNEN

NTNPPQTQTEATGPFSAHEETSRDSDDDPVLIPGARYRAGPGRRSAVA

## CDS

BASE COUNT	670 a	392 c	438 g	524 t	
ORIGIN					
Query Match	61.8%; Score 1594.4; DB 9; Length 2024;				
Best Local Similarity	99.9%; Pred. No. 0;				
Matches 1595; Conservative	0; Mismatches 1; Indels 0; Gaps 0;				
QY	985 CGAAGCAGATGGAGAGAGATCCCAATGTCTATTGATGACAGAAATGCTGATATG 1044				
DB	1 CGAAGCAGATGGAGAGAGATCCCAATGTCTATTGATGACAGAAATGCTGATATG 60				
QY	1045 TTATCAAGATGGTTGAAGAACAGTAGAGTTGCCAAGCAATAGAGAGAGAGAA 1104				
DB	61 TTATCAAGATGGTTGAAGAACAGTAGAGTTGCCAAGCAATAGAGAGAGAGAA 120				
QY	1105 TCTGACCCAGAGTGGAAACAGTCAATCAGATATTTCAACTCTTCTACGGTCCCATCA 1164				
DB	121 TCTGACCCAGAGTGGAAACAGTCAATCAGATATTTCAACTCTTCTACGGTCCCATCA 180				
QY	1165 AGTCCTGATTTGGAAGTGAAGTGAATGCAATGCAATGCAATGCAATGCAATGCAATG 1224				
DB	181 AGTCCTGATTTGGAAGTGAAGTGAATGCAATGCAATGCAATGCAATGCAATGCAATG 240				
QY	1225 CTTACAGCTTCTACATCTCTACATCTCTACATCTCTACATCTCTACATCTCTACATCT 1284				
DB	241 CTTACAGCTTCTACATCTCTACATCTCTACATCTCTACATCTCTACATCTCTACATCT 300				
QY	1285 GAAAGCCCTCATCTTACCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1344				
DB	301 GAAAGCCCTCATCTTACCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 360				
QY	1345 GAGGATCTGGACACACACATCATCAGTCTGATACATCAATATGAAAGTGAAGCCCT 1404				
DB	361 GAGGATCTGGACACACACATCATCAGTCTGATACATCAATATGAAAGTGAAGCCCT 420				
QY	1405 AAACAGGACAGGTGAACAGCTTTTAACTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1464				
DB	421 AAACAGGACAGGTGAACAGCTTTTAACTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 480				
QY	1465 AGCACAATAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1524				
DB	481 AGCACAATAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 540				
QY	1525 ATTGGAGCCATTTGCAATCTGAGGTGAGGAGGATCTTCTGTCACAGAGCTCAGTG 1584				
DB	541 ATTGGAGCCATTTGCAATCTGAGGTGAGGAGGATCTTCTGTCACAGAGCTCAGTG 600				
QY	1585 CAACCCAG 1644				
DB	601 CAACCCAG 660				
QY	1645 AAATATCAGGAGAGATCTGCGAAGAACCCAGTTCAGAACCATATCAATATAACACAA 1704				
DB	661 AAATATCAGGAGAGATCTGCGAAGAACCCAGTTCAGAACCATATCAATATAACACAA 720				
QY	1705 TCAGATAGTTTCAGAGCAAGCCATTTGATTTCCAACTCAGGAGAGAGAGAGAGAGAG 1764				
DB	721 TCAGATAGTTTCAGAGCAAGCCATTTGATTTCCAACTCAGGAGAGAGAGAGAGAGAG 780				
QY	1765 CTTGATCCTCTTTGGGGTTCCAGAGAAATCTGCTTCACTGAAAGCCAGCAAGCAACCA 1824				
DB	781 CTTGATCCTCTTTGGGGTTCCAGAGAAATCTGCTTCACTGAAAGCCAGCAAGCAACCA 840				
QY	1825 GAACTTCAGATCAGACTAGCAGTCTGAGAGTCTGAGAGTCTGAGAGTCTGAGAGTCT 1884				
DB	841 GAACTTCAGATCAGACTAGCAGTCTGAGAGTCTGAGAGTCTGAGAGTCTGAGAGTCT 900				
QY	1885 CCTCAGTTCCAAACAGAGAGCCACTGGGCTTCAGTCTATGAGAACATCCACAGGGAC 1944				

Db	901	CCTCAGTTCACAAACAGAGGCCACTGGGCTTTCAGTCTCATGAGAAACATCCACAGGGAC	960		
QY	1945	TCTGCTCTTCAGGACACAGATGACAGTTCATGACCCAGTCTCTGATCCAGGTCAGAG	2004		
Db	961	TCTGCTCTTCAGGACACAGATGACAGTTCATGACCCAGTCTCTGATCCAGGTCAGAG	1020		
QY	2005	TATCGAGCAGGACCTGGTGTATGACGCTCTGCTGTTGCCGCTATTCAGGATGCTTCAGA	2064		
Db	1021	TATCGAGCAGGACCTGGTGTATGACGCTCTGCTGTTGCCGCTATTCAGGATGCTTCAGA	1080		
QY	2065	CGGAGAAAAGAAAGAAAGAAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGA	2124		
Db	1081	CGGAGAAAAGAAAGAAAGAAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGA	1140		
QY	2125	GTAAAATGGTTTATAAAGGCCATCGCAACTCCAGGACAAATGATAAAGAGCAATTTTC	2184		
Db	1141	GTAAAATGGTTTATAAAGGCCATCGCAACTCCAGGACAAATGATAAAGAGCAATTTTC	1200		
QY	2185	TGGGGTGTCTAACTTTGTAATGAGTGGTCTGACTGTGGCCACATTTTCATCTGGGATCGG	2244		
Db	1201	TGGGGTGTCTAACTTTGTAATGAGTGGTCTGACTGTGGCCACATTTTCATCTGGGATCGG	1260		
QY	2245	CACACTGCTGAGCATTTGATGCTTCTGGAAGCTGATATCATGCTGCTAACTGCTGCTGAG	2304		
Db	1261	CACACTGCTGAGCATTTGATGCTTCTGGAAGCTGATATCATGCTGCTAACTGCTGCTGAG	1320		
QY	2305	CCACATCCGTTTGACCCAAATTTTAGCCTCATCTGGCATAGATTATGACATAAAGATCTGG	2364		
Db	1321	CCACATCCGTTTGACCCAAATTTTAGCCTCATCTGGCATAGATTATGACATAAAGATCTGG	1380		
QY	2365	TCACCATTAAGAGTCAAGGATTTTAAACCGAAACTTGTCTGATGAAGTTTAACTCGA	2424		
Db	1381	TCACCATTAAGAGTCAAGGATTTTAAACCGAAACTTGTCTGATGAAGTTTAACTCGA	1440		
QY	2425	AACGAATCATGCTCGGAAGAACTAGAACACACATTTACAGTTCACGCTTCTTCATGTTG	2484		
Db	1441	AACGAATCATGCTCGGAAGAACTAGAACACACATTTACAGTTCACGCTTCTTCATGTTG	1500		
QY	2485	AGGATGTTGGCTTCACTTAATCATATCCGAGCTGACCGTTGGAGGTTGACAGATCAGAA	2544		
Db	1501	AGGATGTTGGCTTCACTTAATCATATCCGAGCTGACCGTTGGAGGTTGACAGATCAGAA	1560		
QY	2545	GGCTCTGGTCAAGAGATGAATGAGGATGAGAA	2580		
Db	1561	GGCTCTGGTCAAGAGATGAATGAGGATGAGAA	1596		
RESULT 11	HSR01706	1952 bp	mRNA	linear	PRI 20-MAR-2002
LOCUS	HSR01706	cdna DKF2p434C2211 (from clone DKF2p434C2211);			
DEFINITION	Homo sapiens mRNA; complete cds.				
ACCESSION	AL136738	GI:12052994			
VERSION	AL136738.1				
KEYWORDS	human.				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 1952)				
AUTHORS	Duesterhoeft,A., Lauber,J., Mewes,H.W., Gassenhuber,J. and Wiemann,S.				
TITLE	Direct Submission				
JOURNAL	Submitted (12-MAR-2002) MIPS, Am Klopferspitz 18a, D-82152 Martinsried, GERMANY				
COMMENT	Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing consortium of the German Genome Project. This clone (DKF2p434C2211) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzd.de Further information about the clone and the sequencing project is available				



```

FEATURES             at http://mips.gsf.de/prof/cdna/.
    source            Location/Qualifiers
                        1..1952
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /map="lg23.2-24.3"
                        /clone="DKFzp434C2211"
                        /tissue_type="testis"
                        /clone_lib="434 (synonym: htes3). Vector pSport1; host
                        DH10B; sites NotI + SalI"
                        /dev_stage="adult"
                        1..1952
                        /gene="DKFzp434C2211"
                        133..1581
                        /gene="DKFzp434C2211"
                        /note="similarity to mouse PC326"
                        /codon_start=1
                        /product="hypothetical protein"
                        /protein_id="CAB66672.1"
                        /db_xref="GI:12052995"
                        /db_xref="SPTREMBL:O9HOM1"
                        /translation="MEVDTPAEQFLQPTSTSNMSAQHSTSPSTPSTPLSSPDS
                        EORQSVASGHHTHQSDSPSSVYNKQLGMSLDEQDNNNEKLSPKPTGSPVLSLH
                        YSPTGTTTSIKLNFTHSSSVASSRGIGSHCKSEQSEFVPOSSVQPPEDSETK
                        APRESSDVKTYDEGVSAENPVENHNITQSDKF"AKPLDINSGRNDLNDLRSGCVP
                        EGAASDKAKEPSTQSTESATENNNTNPQTEATGPSAHEFTSTRDSALQDT
                        DSDSDVLPJLQYRAGDEGRSAVARIQEYFRFRKKKEKEELDTLNIRPLVKYK
                        YKGRHNSRTMIKEANFAGNFMVMSGDCGHIFINDRHTAHLMLLEADHHVNVCLQPH
                        FPFPIILASSGIDYDIKIWSPSLRSESRIFNKLADYITRNELMLEETRNITVPSAFML
                        RMLASLNHTRADRLLEGDRSEGSQENENDEE"
                        1913..1918
                        /gene="DKFzp434C2211"
                        1932
                        /gene="DKFzp434C2211"
BASE COUNT      642 a   392 c   418 g   500 t
ORIGIN
Query Match      56.1%; Score 1446.4; DB 9; Length 1952;
Best Local Similarity 96.1%; Pred. No. 0;
Matches 1517; Conservative 0; Mismatches 1; Indels 60; Gaps 1;

QY 1063 GAAGCAAGTCAGGTTCACAAAGCAATAGAGCAGCAGGAAGATCTCGACCCAGAGGTGGA 1122
      |||
      1 GAAGCAAGTCAGGTTCGCAAAAGCAATAGAGCAGCAGGAAGATCTCGACCCAGAGGTGGA 60
QY 1123 ACAAGTCAATCAGATATTTCAACTCTTCTCAGGTCCCATCAAGTCCCTGATTGTGAAGTG 1182
      |||
      61 ACAAGTCAATCAGATATTTCAACTCTTCTCAGGTCCCATCAAGTCCCTGATTGTGAAGTG 120
QY 1183 AGTGAACCTGCAATGGAAGTAGATCTCCAGTGAACAATTTCTTCAGCCCTCTACATCC 1242
      |||
      121 AGTGAACCTGCAATGGAAGTAGATCTCCAGTGAACAATTTCTTCAGCCCTCTACATCC 180
QY 1243 TCTACAATGTCAGCTCAGGCTCATTCGACATCATCTCCAGCAGAAAGCCCTCATTTACT 1302
      |||
      181 TCTACAATGTCAGCTCAGGCTCATTCGACATCATCTCCAGCAGAAAGCCCTCATTTACT 240
QY 1303 CTTTGTGCTATCTTCCAGACAGTGAACAAAGGCAGTCTGTGTAGGCATCTGGACACAC 1362
      |||
      241 CTTTGTGCTATCTTCCAGATAGTGAACAAAGGCAGTCTGTGTAGGCATCTGGACACAC 300
QY 1363 ACACATCATCAGTCT- - - - - 1377
      |||
      301 ACACATCATCAGTCTGATTCCTCTCTGTGTTTAAACAACAGTCGATCCATGTCA 360
QY 1378 - - - - - GATAACAATAATGAAAGCTGAGCCCCCAAGCCAGGACAGGTGAA 1422
      |||
      361 CTGTGACGAGCAACGAGTAACAATAATGAAAGCTGAGCCCCCAAGCCAGGACAGTGA 420
QY 1423 CCAGTTTTAAGTTTGCACCTACAGCACAGAAGGAACAACTACAGACCATATAACTGAC 1482
      |||
      421 CCAGTTTTAAGTTTGCACCTACAGCACAGAAGGAACAACTACAGACCAATAAATGAC 480

```

Wed Mar 26 09:38:48 2003

```

|||||
Db 1561 GAAATGAGGATGAGAA 1578

RESULT 12
AK093970
LOCUS
DEFINITION
Homo sapiens cDNA FLJ36651 fis, clone UTERU2000095, highly similar
to Homo sapiens PC326 protein (PC326) mRNA.
ACCESSION
AK093970
VERSION
AK093970.1 GI:21752937
KEYWORDS
oligo capping; fis (full insert sequence).
SOURCE
Homo sapiens uterus cDNA to mRNA, clone_lib:UTERU2
clone:UTERU2000095.

ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1
Ninomiya, K., Nagatsuma, M., Kanda, K., Kondo, H., Yokoi, T.,
Kodaira, H., Furuya, T., Takahashi, M., Kikkawa, E., Omura, Y., Abe, K.,
Kamihara, K., Katsuta, N., Sato, K., Tanikawa, M., Yamazaki, M.,
Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S.,
Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T.,
Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M.,
Kikuchi, H., Murakawa, K., Kanehori, K., Takahashi-Fujii, A.,
Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S.,
Nagahara, K., Masuho, Y., Nagai, K. and Isogai, T.
NEDO human cDNA sequencing project
Unpublished
2
(bases 1 to 2803)
Isogai, T. and Yamamoto, J.
Direct Submission
Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail: genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.

FEATURES
source
1..2803
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UTERU2000095"
/tissue_type="uterus"
/clone_lib="UTERU2"
/note="cloning vector: pME18SFL3"
BASE COUNT 785 a 608 c 639 g 771 t
ORIGIN
Query Match 53.4%; Score 1378.4; DB 9; Length 2803;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1382; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 ATGTCGCGGGTGCTCTACCCACACCTGTTGTTGGGACGTGAGGAAAAGGTCCTCGGG 60
Db 240 ATGTCCTCGGGTGCTCTACCCACACCTGTTGTTGGGACGTGAGGAAAAGGTCCTCGGG 299
QY 61 CTGGAGGACCCGTCCTCGGCTGCGGAGTCGCTACCTGGGAGAGAGAGATTTATCCAAAGA 120
Db 300 CTGGAGGACCCGTCCTCGGCTGCGGAGTCGCTACCTGGGAGAGAGAGATTTATCCAAAGA 359
QY 121 TTAACACTTGAACCAACCTTAATGTCATGATGTTGTGTTAATACAACTCTGTGGAAT 180
Db 360 TTAACACTTGAACCAACCTTAATGTCATGATGTTGTGTTAATACAACTCTGTGGAAT 419
QY 181 GACACTGGAGAAATATATTTAATCTGGTTCAGATGACACCAAAATAGTAATAGTAATCCT 240
Db 420 GACACTGGAGAAATATATTTAATCTGGTTCAGATGACACCAAAATAGTAATAGTAATCCT 479

```



```
repeat_region /rpt_family="L1M4" 12902..13207
repeat_region /rpt_family="AluJo" 13213..13516
repeat_region /rpt_family="AluSx" 13600..13795
repeat_region /rpt_family="L1M4" 13796..14103
repeat_region /rpt_family="AluY" 14104..15110
repeat_region /rpt_family="L1M4" 15111..15412
repeat_region /rpt_family="AluSg" 15413..15877
repeat_region /rpt_family="L1M4" 15867..15963
repeat_region /rpt_family="L1M4" 15953..16049
repeat_region /rpt_family="L1M4" 16050..16766
repeat_region /rpt_family="L1MB8" 16823..17033
repeat_region /rpt_family="MER2" 17034..17080
repeat_region /rpt_family="(TTTTTA)n" complement(17324..17625)
repeat_region /rpt_family="AluSx" complement(18084..18138)
repeat_region /rpt_family="MIR" complement(18566..18706)
repeat_region /rpt_family="MIR" complement(19349..19415)
repeat_region /rpt_family="(CAGAGA)n" complement(20636..21728)
repeat_region /rpt_family="L1MD1" complement(21729..22087)
repeat_region /rpt_family="MLT1A1" complement(22088..22269)
repeat_region /rpt_family="L1MD1" complement(22270..22429)
repeat_region /rpt_family="Alu3b" complement(22488..22840)
repeat_region /rpt_family="L1MD1" 23456..23764
repeat_region /rpt_family="AluY" complement(23767..24338)
repeat_region /rpt_family="L1MD1" 23842
unsure /note="Probably C, possibly A."
repeat_region complement(24635..24942)
repeat_region /rpt_family="AluSg" complement(25255..25293)
repeat_region /rpt_family="L2" 25294..25575
repeat_region /rpt_family="AluSx" complement(25576..25720)
repeat_region /rpt_family="L2" complement(26187..26471)
repeat_region /rpt_family="AluSx" complement(26861..27080)
repeat_region /rpt_family="AluSg/X" complement(27508..28893)
repeat_region /rpt_family="pTR5" 28894..29189
repeat_region /rpt_family="AluSx"

Query Match 49.9%; Score 1287.2; DB 9; Length 162314;
Best Local Similarity 95.8%; Pred. No. 0;
Matches 1322; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 1 ATGCTCTCGGGTGGCTCTCTGCCACATCTGTTGGGACATGAGGAAAAGGTCCTCGGG 60
Db 103593 ATGCTCTCGGGTGGCTCTCTGCCACATCTGTTGGGACATGAGGAAAAGGTCCTCGGG 103534
```



TITLE	Direct Submission
JOURNAL	Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT	NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.
FEATURES	Location/Qualifiers
source	1. .1726 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="HEMBAL002937" /tissue_type="whole embryo, mainly head" /clone_lib="HEMBAL" /dev_stage="embryo, 10 weeks" /note="cloning vector: pME18SFL3" 305 c 365 q 494 t 562 a
sequence	

Query Match	41.4%	Score 1067.4	DB 9	Length 1726
Best Local Similarity	96.8%	Pred. No. 4.5e-280		
Matches 1102	Conservative 0	Mismatches 31	Indels 6	Gaps 1
QY 1442	ACAGCACAGAAGAAACAACTACAAGCACAAATAAAACCTGAACTTTACAGATGATGAGACGA 1501			
DB				
240	AAAGATCATATTTGATTAACCTTCTTGGTGAATCACTTATCTTGTGTTTATGATGATGAGACGA 299			
QY 1502	GTATAGCATCAAGTTCTTAGAGGAATTTGGAGCCATTGCGAATCTGCAATCTGAGGGTCAGGAGGAAT 1561			
DB				
300	GTATAGCATCAAGTTCTTAGAGGAATTTGGAGCCATTGCGAATCTGCAATCTGAGGGTCAGGAGGAAT 359			
QY 1562	CTTTCTGTCACAGAGCTCAGTGCACACCAACCAAGAGGAGACAGTGAACAAAAGCTCCTG 1621			
DB				
360	CTTTCTGTCACAGAGCTCAGTGCACACCAACCAAGAGGAGACAGTGAACAAAAGCTCCTG 419			
QY 1622	AAGATCATCAGAGGATGTGCAAAATATTCAGAAAGGAGTATCTCGAAAAACCCAGTTG 1681			
DB				
420	AAGATCATCAGAGGATGTGCAAAATATTCAGAAAGGAGTATCTCGAAAAACCCAGTTG 479			
QY 1682	AGAACCATATCAATATAACACAAATCAGATTAAGTTCACAGCCAAGCCATTGGATCCCAACT 1741			
DB				
480	AGAACCATATCAATATAACACAAATCAGATTAAGTTCACAGCCAAGCCATTGGATCCCAACT 539			
QY 1742	CAGGAGAAAGAAATGACCTCAATCTTGATGCTCTTGTGGGCTTCAGAAAGAAATCTGCTT 1801			
DB				
540	CAGGAGAAAGAAATGACCTCAATCTTGATGCTCTTGTGGGCTTCAGAAAGAAATCTGCTT 599			
QY 1802	CATCTGAAAAAGCCAAAGGAACCCAGAAACTTCAGATCAGACTAGCACTGAGAGTGCCTACCA 1861			
DB				
600	CATCTGAAAAAGCCAAAGGAACCCAGAAACTTCAGATCAGACTAGCACTGAGAGTGCCTACCA 659			
QY 1862	ATGAAAATAACACCAATCTTGAGGCTCAGTTCACAAACAGAAAGCCACTGGGCCCTTCAGGCTC 1921			
DB				
660	ATGAAAATAACACCAATCTTGAGGCTCAGTTCACAAACAGAAAGCCACTGGGCCCTTCAGGCTC 719			
QY 1922	ATGAGAAACATCCACAGGGACTCTGCTCTTCAGGACACAGATGACAGTGTATGATGACC 1981			
DB				
720	ATGAGAAACATCCACAGGGACTCTGCTCTTCAGGACACAGATGACAGTGTATGATGACC 779			
QY 1982	CAGTCTGTATCCCGGTCAGGTTATCGACAGGACCTGGTGATAGACGCTCTGCTGTGTG 2041			
DB				
780	CAGTCTGTATCCCGGTCAGGTTATCGACAGGACCTGGTGATAGACGCTCTGCTGTGTG 839			
QY 2042	CCGCTATTCCAGGATTTCTTCAGCGGGAGAAAGAAAGAAATGGAAATTTGGATA 2101			
DB				
840	CCGCTATTCCAGGATTTCTTCAGCGGGAGAAAGAAAGAAATGGAAATTTGGATA 899			
QY 2102	CTTTTGAACATTTAGAGGGCGCTAGTTAAATATGGTTTATTAAGGGCCATTCGCAACTCCAGGA 2161			
DB				

Db 900 CTTTGAACATTAGAAGCCCGCTAGTAAATGTTTATAAAGCCATCCCACTCCAGGA 959

QY 2162 CAATGATAAAGAGCAATTTCTGGGTCCTAACTTTGTAATGAGTGGTTCFCACTGTG 2221

Db 960 CAATGATAAAGAGCAATTTCTGGGTCCTAACTTTGTAATGAGTGGTTCFCACTGTG 1019

QY 2222 GCACATTTTCATCTGCGGATCGGCACACCTGCTGAGCATTTGATCTCTGGAAGCTGATA 2281

Db 1020 GCACATTTTCATCTGCGGATCGGCACACCTGCTGAGCATTTGATCTCTGGAAGCTGATA 1079

QY 2282 ATCATGTGTGTAACCTGCTGAGCCACATCCGTTTGACCAATTTTATGCTCATCTGGCA 2341

Db 1080 ATCATGTGTGTAACCTGCTGAGCCACATCCGTTTGACCAATTTTATGCTCATCTGGCA 1133

QY 2342 TAGATTATGACATAAAGATCTGGTCACCATTAGAAGTCAAGGATTTTAAACCAAAAC 2401

Db 1134 TAGATTATGACATAAAGATCTGGTCACCATTAGAAGTCAAGGATTTTAAACCAAAAC 1193

QY 2402 TTGCTGATGAAGTTATACCTGAAACGAACATGCTGCGAAGAACTAGAACACCATTA 2461

Db 1194 TTGCTGATGAAGTTATACCTGAAACGAACATGCTGCGAAGAACTAGAACACCATTA 1253

QY 2462 CAGTTCAGCCTCTTCATGTTGAGGATGTTGGCTTACCTTAATCATATCCGACTGACC 2521

Db 1254 CAGTTCAGCCTCTTCATGTTGAGGATGTTGGCTTACCTTAATCATATCCGACTGACC 1313

QY 2522 GGTGGAGGTGACAGATCAGAGGCTCTGGTCAAGAGATGAAATGAGGATGAGGAA 2580

Db 1314 GGTGGAGGTGACAGATCAGAGGCTCTGGTCAAGAGATGAAATGAGGATGAGGAA 1372

RESULT 16

AC107981

LOCUS

DEFINITION Homo sapiens chromosome 15 clone CTD-3247F8 map 15, LOW-PASS

ACCESSION AC107981

VERSION

KEYWORDS

SOURCE HTG; HTGS; PHASE0.

ORGANISM Homo sapiens

REFERENCE

1 (bases 1 to 65517)

1. Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, A., Brown, A., Cantata, J., Campolano, A., Chang, J., Chazaro, B., Choquet, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DeArillano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliiev, I., Johnson, R., Jones, C., Kamat, A., Karas, A., Kellis, C., Larocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T., Mienda, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tropham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

TITLE

JOURNAL

Research, 320 Charles Street) Whitehead Institute/MIT Center for Genome

## COMMENT

All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information  
Center project name: L24536  
Center clone name: 3247\_F\_8

\* NOTE: This record contains 84 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 665: contig of 665 bp in length  
766 765: gap of 100 bp  
1461 1560: contig of 695 bp in length  
1561 2254: contig of 694 bp in length  
2255 2354: gap of 100 bp  
2355 3019: contig of 665 bp in length  
3020 3119: gap of 100 bp  
3120 3761: contig of 642 bp in length  
3762 3861: gap of 100 bp  
3862 4525: contig of 664 bp in length  
4526 4625: gap of 100 bp  
4626 5280: contig of 655 bp in length  
5281 5380: gap of 100 bp  
5381 6092: contig of 712 bp in length  
6093 6192: gap of 100 bp  
6193 6884: contig of 692 bp in length  
6885 6984: gap of 100 bp  
6985 7703: contig of 719 bp in length  
7704 7803: gap of 100 bp  
7804 8468: contig of 665 bp in length  
8469 8568: gap of 100 bp  
8569 9249: contig of 681 bp in length  
9250 9349: gap of 100 bp  
9350 10034: contig of 685 bp in length  
10035 10134: gap of 100 bp  
10135 10797: contig of 663 bp in length  
10798 10897: gap of 100 bp  
10898 11584: contig of 687 bp in length  
11585 11684: gap of 100 bp  
11685 12381: contig of 697 bp in length  
12382 12481: gap of 100 bp  
12482 13169: contig of 688 bp in length  
13170 13269: gap of 100 bp  
13270 13973: contig of 704 bp in length  
13974 14073: gap of 100 bp  
14074 14774: contig of 701 bp in length  
14775 14874: gap of 100 bp  
14875 15569: contig of 695 bp in length  
15570 15669: gap of 100 bp  
15670 16331: contig of 662 bp in length  
16332 16431: gap of 100 bp  
16432 17142: contig of 711 bp in length  
17143 17242: gap of 100 bp  
17243 17916: contig of 674 bp in length  
17917 18016: gap of 100 bp  
18017 18704: contig of 688 bp in length  
18705 18804: gap of 100 bp  
18805 19462: contig of 658 bp in length  
19463 19562: gap of 100 bp









[illegible]



such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1>

703H14 is from the library RPC14 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see [http://bacpac.med.buffalo.edu/ VECTOR:pcvPAC2](http://bacpac.med.buffalo.edu/VECTOR:pcvPAC2)

IMPORTANT: This sequence is not the entire insert of clone 703H14. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

The true left end of clone 702J19 is at 67710 in this sequence. The true right end of clone 295C6 is at 100 in this sequence.

## FEATURES

## source

```
1. .74729
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="q23.2-24.3"
/clone="RP4-703H14"
/clone_lib="RPC1-4"
```

```
5. .112
/note="L1M2 repeat: matches 1519. .1625 of consensus"
457. .754
/note="AlusX repeat: matches 2. .299 of consensus"
824. .935
/note="L2 repeat: matches 2629. .2750 of consensus"
1050. .1570
/note="L1M40a repeat: matches 3. .519 of consensus"
2795. .3025
/note="L2 repeat: matches 2492. .2750 of consensus"
3038. .3096
/note="MIR repeat: matches 192. .242 of consensus"
3097. .3399
/note="AlusX repeat: matches 1. .303 of consensus"
3400. .3553
/note="MIR repeat: matches 3. .192 of consensus"
Join(3604. .3665,18433. .18525,26634. .26819,39299. .39412,
43008. .43143,45030. .45244,54286. .54379,55674. .55793,
56337. .56597,67585. .67763)
/feature="dJ703H14.1"
Join(<3604. .3665,18433. .18525,26634. .26819,39299. .39412,
43008. .43143,45030. .45244,54286. .54379,55674. .55793,
56337. .56597,67585. .67763)
/feature="dJ703H14.1"
/feature="match: ESTs: Em:N54388 Em:RL4260 Em:H08089
Em:Z24943 Em:AAL30053"
/codon_start=3
/evidence=not_experimental
/protein_id="CAB52430.1"
/db_xref="GI:5725427"
/translation="RRRIQRLKLEATLNVHDCVNTICWMDTGYILSGSDTKLVI
SNPKGRVLTIRSGHRANIPFAKELPCTNDKQIVSCSGDGVIFVNEQDAETNRQC
QFTCHGYTTEITVPNPYFLSCGEDGTWFTDRIKTSCTKEDCKDDILLNCRA
ASVAICPIPYLLAVGCSDSVRIYDRMLGTRATGNAGTGTGMVARTIPSLNN
KSCRTSLCYSDGQILVSYSDYIYLPDKDDTAREIKTFSAEERREELRQPPVKR
LRLCDWSDTGPRAPEPRDQSPNVSLMRMSDMLSRWFEEAEVAQSNRGRG
SPRPSGTSQSDISLTLPVSPDLEVSTAMEVDTPAQFLOPSTSTSMASQAHSST
SPSTPSTPLTLSSSDSEQRQSVSEAGHHTHQSEFLRGPEIALLKRLLQRLKRAE
QRRQELAAHTQQPSTSDQSSHEGSSQDPHAGY"
```

```
repeat_region
/feature="L2 repeat: matches 1292. .1395 of consensus"
3993. .4316
/feature="Alu repeat: matches 4. .311 of consensus"
4317. .4657
/feature="L2 repeat: matches 1395. .1804 of consensus"
complement(4658. .4961)
/feature="match: GSS: Em:AQ488382.1"
/evidence=not_experimental
```

```
repeat_region
/feature="L2 repeat: matches 2323. .2621 of consensus"
5417. .5702
/feature="AlusX repeat: matches 1. .288 of consensus"
5710. .5804
/feature="HALI repeat: matches 734. .826 of consensus"
5810. .6092
/feature="AlusX repeat: matches 5. .301 of consensus"
6245. .6554
/feature="AluY repeat: matches 2. .311 of consensus"
6571. .7311
/feature="L1M1 repeat: matches 5415. .6318 of consensus"
7319. .7583
/feature="L1M4 repeat: matches 4053. .4324 of consensus"
7584. .7869
/feature="AluJo repeat: matches 1. .281 of consensus"
7870. .8013
/feature="L1M4 repeat: matches 3912. .4053 of consensus"
8010. .8842
/feature="L1M4 repeat: matches 2663. .3517 of consensus"
8917. .10364
/feature="L1M1 repeat: matches 4735. .6165 of consensus"
10529. .12088
/feature="L1M1 repeat: matches 4735. .6316 of consensus"
12098. .12385
/feature="AlusX repeat: matches 21. .303 of consensus"
12386. .12422
/feature="dJ703H14.1"
/feature="match: GSS: Em:AQ081840.1"
/evidence=not_experimental
12386. .12420
/feature="dJ703H14.1"
/feature="match: GSS: Em:AQ265659"
/evidence=not_experimental
complement(12387. .12420)
/feature="match: GSS: Em:AQ000534
match: GSS: Em:AQ001615"
/evidence=not_experimental
12426. .12589
/feature="L1M4 repeat: matches 2686. .2851 of consensus"
12590. .13036
/feature="L1M1 repeat: matches 4265. .4711 of consensus"
13037. .13344
/feature="AluY repeat: matches 1. .309 of consensus"
13345. .13542
/feature="L1M1 repeat: matches 4068. .4265 of consensus"
13546. .13630
/feature="trigger4(2ombi) repeat: matches 44. .146 of consensus"
13631. .13646
/feature="MER46A repeat: matches 2178. .3011 of consensus"
13650. .14477
/feature="L1MEC repeat: matches 1. .302 of consensus"
14478. .14778
/feature="AlusX repeat: matches 2. .293 of consensus"
14783. .15077
/feature="AlusX repeat: matches 2403. .2179 of consensus"
15078. .15206
/feature="L1MEC repeat: matches 3. .312 of consensus"
15230. .15530
/feature="AlusX repeat: matches 581. .2130 of consensus"
15837. .17435
/feature="L1M4 repeat: matches 1. .296 of consensus"
17436. .17738
/feature="AlusX repeat: matches 248. .581 of consensus"
17739. .18061
/feature="L1M4 repeat: matches 2. .230 of consensus"
18086. .18306
/feature="MER30 repeat: matches 18528
18307. .18528
/feature="dJ703H14.1"
/feature="match: GSS: Em:AQ083937"
/evidence=not_experimental
19519. .19834
repeat_region
```

Contact: Thomas Hudson  
Whitehead Institute/MIT Center for Genome Research  
Whitehead Institute for Biomedical Research  
9 Cambridge Center, Cambridge MA 02142 USA

RESULT 21	AC069147	158324 bp	DNA	linear	HTG 01-SEP-2000
LOCUS	AC069147				
DEFINITION	Homo sapiens chromosome 1 clone RP11-105H9, WORKING DRAFT SEQUENCE,				
	AC069147	7 unordered pieces.			
ACCESSION	AC069147				
VERSION	AC069147.3	GI:8099949			
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.				

Wed Mar 26 09:38:48 2003

```

SOURCE          Homo sapiens.
ORGANISM        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1 (bases 1 to 158324)
                Waterston, R.H.
TITLE          The sequence of Homo sapiens clone
                Unpublished
REFERENCE       2 (bases 1 to 158324)
                Waterston, R.H.
AUTHORS        Direct Submission
TITLE          Submitted (19-MAY-2000) Genome Sequencing Center, Washington
JOURNAL        University School of Medicine, 4444 Forest Park Parkway, St. Louis,
                MO 63108, USA
COMMENT        On May 28, 2000 this sequence version replaced gi:8028173.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0105H09
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 154865 bases at least Q40
Consensus quality: 155623 bases at least Q30
Consensus quality: 156154 bases at least Q20
Insert size: 174000; agarose-fp
Insert size: 157724; sum-of-contigs
Quality coverage: 7.55 in Q20 bases; agarose-fp
Quality coverage: 8.39 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 6587: contig of 6587 bp in length
* 6588 6687: gap of unknown length
* 6688 14247: contig of 7560 bp in length
* 14248 14347: gap of unknown length
* 14348 28850: contig of 14503 bp in length
* 28851 49498: gap of unknown length
* 49499 49598: contig of 20548 bp in length
* 49599 69546: gap of unknown length
* 69547 69646: contig of 19948 bp in length
* 69647 105850: gap of unknown length
* 105851 105951: contig of 36204 bp in length
* 105951 158324: contig of 52374 bp in length.
*
* Location/Qualifiers
* 1. 158324
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="11"
* /clone="RP11-105H9"
*
* 1. 6587
* /note="assembly_name:Contig15"
* 6688. 14247
* /note="assembly_name:Contig16"
* 14348. 28850
* /note="assembly_name:Contig17"
* /note="assembly_name:Contig17"
* clone_end:77
* vector_side:right
* 28951. 49498
* /note="assembly_name:Contig18"

```

```

misc_feature    49599..69546
                /note="assembly_name:Contig19"
misc_feature    69647..105850
                /note="assembly_name:Contig20"
misc_feature    105951..158324
                /note="assembly_name:Contig21"
                clone_end:SP6
                vector_side:right
BASE COUNT      47660 a 30780 c 31018 g 48209 t      657 others
ORIGIN
Query Match      8.4%; Score 217.4; DB 2; Length 158324;
Query Local Similarity 99.5%; Pred. No. 9.9e-48;
Matches 218; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 687 AGGGATTATGACGGTCGAGGAGGACTACTGGAATGGTGGCCGTTTATTCCTTCCCATCT 746
      |||||||
DB 25723 AGGGAATTATGACGGTCGAGGAGGACTACTGGAATGGTGGCCGTTTATTCCTTCCCATCT 25782
      |||||||
QY 747 TAATAATAAGTCTGTCAGAGTGACATCTCTGTGTACAGTGAAGATGGTCAAGAGATTCT 806
      |||||||
DB 25783 TAATAATAAGTCTGTCAGAGTGACATCTCTGTGTACAGTGAAGATGGTCAAGAGATTCT 25842
      |||||||
QY 807 CGTTAGTACTCTTCAGATTACATATATCTTTTGTACCCGAAAGATGATACAGCAGGAGA 866
      |||||||
DB 25843 CGTTAGTACTCTTCAGATTACATATATCTTTTGTACCCGAAAGATGATACAGCAGGAGA 25902
      |||||||
QY 867 ACTTAAACTCTCTTCGCGAAGAGAGAGAGAGAGTT 905
      |||||||
DB 25903 ACTTAAACTCTCTTCGCGAAGAGAGAGAGAGAGTT 25941
      |||||||

RESULT 22
AC026140/c      163230 bp      DNA      linear      HTG 26-MAY-2000
LOCUS           Homo sapiens chromosome 1 clone RP11-300C7 map 1, WORKING DRAFT
DEFINITION      SEQUENCE, 17 unordered pieces.
ACCESSION       AC026140
VERSION         AC026140.4 GI:8077102
KEYWORDS        HTG; HTGS; PHASE1; HTGS_DRAFT.
SOURCE          Homo sapiens.
ORGANISM        Homo sapiens
REFERENCE        1 (bases 1 to 163230)
AUTHORS          Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE           Homo sapiens chromosome 1, clone RP11-300C7
JOURNAL         Unpublished
AUTHORS          2 (bases 1 to 163230)
                Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
                Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bada, F.,
                Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G.,
                Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
                Collimore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S.,
                Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
                Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
                Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
                Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
                Klein, J., Laroque, K., Lamazares, R., Landers, T., Lehoczy, J.,
                Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, J.,
                McCarthy, M., McEwan, P., McKernan, K., Miranda, C., Mieng, V., Morrow, J.,
                Meldrim, J., Meneus, L., Mihova, T., Miranda, C. H., O'Connor, T., O'Donnell, P.,
                Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
                O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,
                Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
                Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
                Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
                Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
                Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
                Young, G., Zainoun, J., Zimmer, A. and Zody, M.
                Direct Submission
                Submitted (19-MAR-2000) Whitehead Institute/MIT Center for Genome
                Research, 320 Charles Street, Cambridge, MA 02141, USA

```

## COMMENT

On May 25, 2000 this sequence version replaced gi:7523800.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L8103

Center clone name: 300\_C\_7

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 152621 bases at least Q40

Consensus quality: 157889 bases at least Q30

Consensus quality: 159998 bases at least Q20

Insert size: 157000; agarose-fp

Insert size: 161630; sum-of-contigs

Quality coverage: 4.9 in Q20 bases; agarose-fp

Quality coverage: 4.8 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 17 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence,  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

1 1941: contig of 1941 bp in length
2 1942 2041: gap of 100 bp
3 2042 3714: contig of 1673 bp in length
4 3715 3814: gap of 100 bp
5 3815 7590: contig of 3776 bp in length
6 7591 7690: gap of 100 bp
7 7691 11450: contig of 3760 bp in length
8 11451 11550: gap of 100 bp
9 11551 18921: contig of 7371 bp in length
10 18922 19021: gap of 100 bp
11 19022 25215: contig of 6194 bp in length
12 25216 25315: gap of 100 bp
13 25316 30271: contig of 4956 bp in length
14 30272 30371: gap of 100 bp
15 30372 38111: contig of 7740 bp in length
16 38112 38211: gap of 100 bp
17 38212 47808: contig of 9597 bp in length
18 47809 47908: gap of 100 bp
19 47909 58047: contig of 10139 bp in length
20 58048 58147: gap of 100 bp
21 58148 71836: contig of 13689 bp in length
22 71837 71936: gap of 100 bp
23 71937 81124: contig of 9188 bp in length
24 81125 81224: gap of 100 bp
25 81225 92003: contig of 10778 bp in length
26 92003 92102: gap of 100 bp
27 92103 105933: contig of 13831 bp in length
28 105934 106033: gap of 100 bp
29 106034 121911: contig of 15878 bp in length
30 121912 122011: gap of 100 bp
31 122012 141942: contig of 19931 bp in length
32 141943 142042: gap of 100 bp
33 142043 163230: contig of 21188 bp in length.
    Location/Qualifiers
      1. .163230
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="1"
        /map="1"
        /clone="RP11-300C7"
        /clone_lib="RPC1-11 Human Male BAC"

```

FEATURES  
source

```

misc_feature 1. .1941
              /note="assembly_fragment"
misc_feature 2042. .3714
              /note="assembly_fragment"
              clone_end:77
              vector_side:right
misc_feature 3815. .7590
              /note="assembly_fragment"
misc_feature 7691. .11450
              /note="assembly_fragment"
misc_feature 11551. .18921
              /note="assembly_fragment"
misc_feature 19022. .25215
              /note="assembly_fragment"
misc_feature 25316. .30271
              /note="assembly_fragment"
misc_feature 30372. .38111
              /note="assembly_fragment"
misc_feature 38212. .47808
              /note="assembly_fragment"
misc_feature 47909. .58047
              /note="assembly_fragment"
misc_feature 58148. .71836
              /note="assembly_fragment"
misc_feature 71937. .81124
              /note="assembly_fragment"
misc_feature 81225. .92002
              /note="assembly_fragment"
misc_feature 92103. .105933
              /note="assembly_fragment"
misc_feature 106034. .121911
              /note="assembly_fragment"
misc_feature 122012. .141942
              /note="assembly_fragment"
misc_feature 142043. .163230
              /note="assembly_fragment"
              clone_end:SP6
              vector_side:right

BASE COUNT 52925 a 30771 c 29634 g 48296 t 1604 others
ORIGIN

Query Match      8.4%; Score 217.4; DB 2; Length 163230;
Best Local Similarity 99.5%; Pred. No. 9.9e-48;
Matches 218; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 687 AGGGAATTATGAGGTCGAGGACTACTGGAATGGTTGCCGTTTATTCCTCCCATCT 746
      |||||
Db 97524 AGGGAATTATGAGGTCGAGGACTACTGGAATGGTTGCCGTTTATTCCTCCCATCT 7465

QY 747 TAATAAAGTCCTGCAGAGTGCACATCTCTGTGTACAGTGAAGATGGTCAAGAGATTCT 806
      |||||
Db 97464 TAATAAAGTCCTGCAGAGTGCACATCTCTGTGTACAGTGAAGATGGTCAAGAGATTCT 97405

QY 807 CGTTAGTTACTCTTCAGATTACATATATCTTTTACCCGAAAGATGATACAGCAGAGA 866
      |||||
Db 97404 CTTTACTCTCTTCAGATTACATATATCTTTTACCCGAAAGATGATACAGCAGAGA 97345

QY 867 ACTTAAACTCTCTTCGCGAAGAGAGAGAGAGAGAGTT 905
      |||||
Db 97344 ACTTAAACTCTCTTCGCGAAGAGAGAGAGAGAGTT 97306

RESULT 23
AC130131/c
LOCUS
DEFINITION
      Rattus norvegicus clone CH230-112N9, *** SEQUENCING IN PROGRESS
      ***, 60 unordered pieces.
ACCESSION
      AC130131
VERSION
      AC130131.1 GI:22138359
KEYWORDS
      HTG; HTGS_PHASE1.
SOURCE
      Rattus norvegicus
      ORGANISM
        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

\* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1152: contig of 1152 bp in length  
 1153: gap of unknown length  
 2369: contig of 1117 bp in length  
 2469: gap of unknown length  
 3720: contig of 1151 bp in length  
 3621: gap of unknown length  
 4891: contig of 1171 bp in length  
 4991: gap of unknown length  
 6504: contig of 1513 bp in length  
 6604: gap of unknown length  
 7660: contig of 1056 bp in length  
 7661: gap of unknown length  
 9156: contig of 1396 bp in length  
 9256: gap of unknown length  
 10289: contig of 1033 bp in length  
 10389: gap of unknown length  
 11437: contig of 1048 bp in length  
 11537: gap of unknown length  
 13109: contig of 1572 bp in length  
 13209: gap of unknown length  
 14230: contig of 1021 bp in length  
 14330: gap of unknown length  
 15551: contig of 1321 bp in length  
 15751: gap of unknown length  
 17898: contig of 2147 bp in length  
 17998: gap of unknown length  
 19308: contig of 1310 bp in length  
 19408: gap of unknown length  
 20992: contig of 1584 bp in length  
 21092: gap of unknown length  
 22726: contig of 1634 bp in length  
 22826: gap of unknown length  
 23869: contig of 1043 bp in length  
 23969: gap of unknown length  
 25600: contig of 1631 bp in length  
 25601: gap of unknown length  
 27123: contig of 1423 bp in length  
 27223: gap of unknown length  
 28233: contig of 1010 bp in length  
 28333: gap of unknown length  
 29780: contig of 1447 bp in length  
 29880: gap of unknown length  
 31396: contig of 1516 bp in length  
 31496: gap of unknown length  
 32754: contig of 1258 bp in length  
 32854: gap of unknown length  
 35174: contig of 2320 bp in length  
 35274: gap of unknown length  
 37272: contig of 1998 bp in length  
 37372: gap of unknown length  
 39148: contig of 1776 bp in length  
 39248: gap of unknown length  
 40423: contig of 1175 bp in length  
 40523: gap of unknown length  
 41917: contig of 1394 bp in length  
 42017: gap of unknown length  
 45246: contig of 3229 bp in length  
 45346: gap of unknown length  
 47170: contig of 1824 bp in length  
 47270: gap of unknown length  
 49917: contig of 2647 bp in length  
 50017: gap of unknown length  
 51247: contig of 1230 bp in length  
 51347: gap of unknown length  
 54015: contig of 2668 bp in length  
 54115: gap of unknown length  
 56300: contig of 2185 bp in length  
 56301: gap of unknown length  
 58739: contig of 2339 bp in length

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

1 (bases 1 to 168307)  
 Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,  
 Alsbrooks, S.L., Amarutunge, H.C., Are, J.R., Ayale, M., Banks, T.,  
 Barberia, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D.,  
 Bouck, J., Bowie, S., Brieva, M., Brown, M., Brown, M., Bryant, N.P.,  
 Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,  
 Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
 Chen, G., Chen, K., Chen, Z., Chowdhry, I., Christopoulos, C.,  
 Cleveland, C.D., Cox, C., Coyle, M.D., Bathorne, S.R., David, R.,  
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,  
 Delanaye, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,  
 Douthett, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
 Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,  
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,  
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
 Harris, C., Harris, K., Hart, M., Haviak, P., Hayes, A., Hernandez, J.,  
 Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.E.,  
 Homsif, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,  
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudan, S.,  
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,  
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,  
 Li, J., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, W., Louseged, H.,  
 Lozada, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,  
 Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,  
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,  
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
 Nguyen, N., Nickerson, E., Nwokenwo, S., Ogih, M., Okwuonu, G.,  
 Oreguene, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, X.,  
 Rivers, M., Rojas, A., Rojokan, I., Rolfe, M., Ruiz, S., Savery, G.,  
 Scherer, S., Scott, G., Shen, H., Shoshitari, N., Sison, I.,  
 Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,  
 Sutton, A., Syatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
 Usmani, K., Vasquez, L., Vera, V., Villalobos, D., Vinson, R., Wang, Q.,  
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
 Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K.,  
 Wu, C., Wu, Y.F., Zhou, J., Zorrilla, S., Zorrilla, S., Nelson, D.,  
 Weinstein, G., and Gibbs, R.

Direct Submission  
 2 (bases 1 to 168307)  
 Worley, K.C.  
 Submitted (08-AUG-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

----- Genome Center  
 ----- Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GHZO  
 Center clone name: CH230-112N9  
 ----- Summary Statistics  
 Sequencing vector: Plasmid  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 116575 bases at least Q40  
 Consensus quality: 125466 bases at least Q30  
 Consensus quality: 132182 bases at least Q20

\* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 60 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT



```

* 58740 58839: gap of unknown length
* 58840 60887: contig of 2048 bp in length
* 60888 60987: gap of unknown length
* 60988 63164: contig of 2177 bp in length
* 63165 63264: gap of unknown length
* 63265 66031: contig of 2767 bp in length
* 66032 66131: gap of unknown length
* 66132 69106: contig of 2975 bp in length
* 69107 69206: gap of unknown length
* 69207 70782: contig of 1576 bp in length
* 70783 70882: gap of unknown length
* 70883 73330: contig of 2448 bp in length
* 73331 73430: gap of unknown length
* 73431 77654: contig of 4224 bp in length
* 77655 77754: gap of unknown length
* 77755 80207: contig of 2452 bp in length
* 80208 80306: gap of unknown length
* 80307 83534: contig of 3228 bp in length
* 83535 83635: gap of unknown length
* 83635 86725: contig of 3091 bp in length
* 86726 86825: gap of unknown length
* 86826 89962: contig of 3137 bp in length
* 89963 90062: gap of unknown length
* 90063 94214: contig of 4152 bp in length
* 94215 94314: gap of unknown length
* 94315 96827: contig of 2513 bp in length
* 96828 96927: gap of unknown length
* 96928 100859: contig of 3932 bp in length
* 100860 100959: gap of unknown length
* 100960 106168: contig of 5209 bp in length
* 106169 106269: gap of unknown length
* 106269 109396: contig of 3127 bp in length
* 109396 114724: contig of 5229 bp in length
* 114725 114824: gap of unknown length
* 114825 119043: contig of 4219 bp in length
* 119044 124368: contig of 5225 bp in length
* 124369 124469: gap of unknown length
* 124469 130352: contig of 5884 bp in length
* 130353 130452: gap of unknown length
* 130453 135340: contig of 4888 bp in length
* 135341 135441: gap of unknown length
* 135441 142967: contig of 7527 bp in length
* 142968 143067: gap of unknown length

Query Match
Best Local Similarity 7.9%; Score 202.8; DB 2; Length 168307;
Matches 259; Conservative 0; Mismatches 67; Indels 6; Gaps 1;

QY 1696 ATACACAACTCAGATAGTTCCACAGCCAGCCATTGGATTCCAACTCAGGAGAAAGAAAT 1755
Db 142896 ACAGCAAACTTGATTAATTTCCCCACATGAGCCATTGGATTCTAACTCAGGAGAAAGAAAT 142837

QY 1756 GACCTCAATCTTGATCGCTCTTGTGGGGTTCACAGAAAGATCGTCTATCTGTAAGAACCC 1815
Db 142836 AACCCAGTCTTGACAGCCCTTGTGGGGTTCACAGAAAGATCGTCTGTCGAGACACAC 142777

QY 1816 AAGGAACCAAGAACTTCAGATCAGATAGCAGTACAGTGTCTACCAATGAAATTAACACC 1875
Db 142776 AGG-----GAGACTTGTGAGCAGACAGCAGTACAGTGTCTACCGAGCGTGCAGCGAC 142723

QY 1876 AATCCTGAGCCTCAGTTCCAAAGAGAGCCAGCCAGTGGCTTCAGTCTATGTAAGAACATCC 1935
Db 142722 AATCCTGAGTCTCCATCCAGACAGAGCCATAGCCCTTCAGTCTATGTAAGAACATCC 142663

QY 1936 ACCAGGACTCTGCTCTTCAGGACACAGATGACAGTGTGATGATGACCCAGTCCCTGATCCCA 1995
Db 142662 GCAGGAGTCTGCTCTTCAGGACACAGATGACAGCGATGATGACCCGCTCTGATCCCC 142603

QY 1996 GTGCAAGGTATCGACGAGCACTGTGTATAG 2027
Db 142602 GTGCAAGATACCGACGAGCACTGTGTATAG 142571

```

## RESULT 24

AC116374

LOCUS

DEFINITION

AC116374

AC116374

HTG: HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.

Mus musculus

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 179269)

Mus musculus chromosome, clone RP23-29419

Unpublished

2 (bases 1 to 179269)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N.,

Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,

Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,

Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collamore, A.,

Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,

Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,

Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,

Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,

Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R.,

Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G.,

MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,

McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L.,

Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,

Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,

Oliver, J., Peterson, K., Phukhang, P., Pierre, N., Pollara, V.,

Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,

Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R.,

Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,

Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,

Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,

Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,

Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (27-MAR-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 179269)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N.,

Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,

Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,

Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collamore, A.,

Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,

Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,

Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,

Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,

Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R.,

Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G.,

MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,

McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L.,

Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,

Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,

Oliver, J., Peterson, K., Phukhang, P., Pierre, N., Pollara, V.,

Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,

Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R.,

Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,

Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,

Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,

Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,

Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (28-MAR-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Mar 28, 2002 this sequence version replaced gi:19745074.

All repeats were identified using RepeatMasker:

Smit, A.F.A. &amp; Green, P. (1996-1997)

TITLE

JOURNAL

AUTHORS

REFERENCE

COMMENT

http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L19554  
 Center clone name: 294\_I\_9  
 ----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 172079 bases at least Q40  
 Consensus quality: 175437 bases at least Q30  
 Consensus quality: 176346 bases at least Q20  
 Insert size: 176000; agarose-fp  
 Insert size: 176969; sum-of-contigs  
 Quality coverage: 8.0 in Q20 bases; agarose-fp  
 Quality coverage: 8.0 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently  
 consists of 24 contigs. Gaps between the contigs  
 are represented as runs of N. The order of the pieces  
 is believed to be correct as given, however the sizes  
 of the gaps between them are based on estimates that have  
 been provided by the submitter.  
 This sequence will be replaced  
 by the finished sequence as soon as it is available and  
 the accession number will be preserved.

1 772: contig of 772 bp in length  
 773 872: gap of 100 bp  
 873 2734: contig of 1862 bp in length  
 2735 2834: gap of 100 bp  
 2835 4210: contig of 1376 bp in length  
 4211 4310: gap of 100 bp  
 4311 5637: contig of 1327 bp in length  
 5638 5737: gap of 100 bp  
 5738 7170: contig of 1433 bp in length  
 7171 7270: gap of 100 bp  
 7271 8657: contig of 1387 bp in length  
 8658 8757: gap of 100 bp  
 8758 10043: contig of 1286 bp in length  
 10044 10143: gap of 100 bp  
 10144 11873: contig of 1730 bp in length  
 11874 11973: gap of 100 bp  
 11974 13309: contig of 1336 bp in length  
 13310 13409: gap of 100 bp  
 13410 14723: contig of 1314 bp in length  
 14724 14823: gap of 100 bp  
 14824 16735: contig of 1912 bp in length  
 16736 16835: gap of 100 bp  
 16836 19688: contig of 2853 bp in length  
 19689 19788: gap of 100 bp  
 19789 22930: contig of 3142 bp in length  
 22931 23030: gap of 100 bp  
 23031 25546: contig of 2516 bp in length  
 25547 25646: gap of 100 bp  
 25647 28158: contig of 2512 bp in length  
 28159 28258: gap of 100 bp  
 28259 32365: contig of 4107 bp in length  
 32366 32465: gap of 100 bp  
 32466 36376: contig of 3911 bp in length  
 36377 36476: gap of 100 bp  
 36477 40247: contig of 3771 bp in length  
 40248 40347: gap of 100 bp  
 40348 82256: contig of 41909 bp in length  
 82257 82357: gap of 100 bp  
 82358 92140: contig of 9784 bp in length  
 92141 92240: gap of 100 bp  
 92241 108763: contig of 16523 bp in length  
 108764 108863: gap of 100 bp  
 108864 141973: contig of 33110 bp in length

\* 141974 142073: gap of 100 bp  
 \* 142074 170270: contig of 28197 bp in length  
 \* 170271 170370: gap of 100 bp  
 \* 170371 179269: contig of 8899 bp in length.

FEATURES	Source
misc_feature	1. 179269 /organism="Mus musculus" /db_xref="taxon:10090" /clone_lib="RPCI-23 Female Mouse BAC"
misc_feature	1. 772 /note="assembly_fragment"
misc_feature	873. 2734 /note="assembly_fragment"
misc_feature	2835. 4210 /note="assembly_fragment"
misc_feature	4311. 5637 /note="assembly_fragment"
misc_feature	5738. 7170 /note="assembly_fragment"
misc_feature	7271. 8657 /note="assembly_fragment"
misc_feature	8758. 10043 /note="assembly_fragment"
misc_feature	10144. 11873 /note="assembly_fragment"
misc_feature	11974. 13309 /note="assembly_fragment"
misc_feature	13410. 14723 /note="assembly_fragment"
misc_feature	14824. 16735 /note="assembly_fragment"
misc_feature	16836. 19688 /note="assembly_fragment"
misc_feature	19789. 22930 /note="assembly_fragment"
misc_feature	23031. 25546 /note="assembly_fragment"
misc_feature	25647. 28158 /note="assembly_fragment"
misc_feature	28259. 32365 /note="assembly_fragment"
misc_feature	32466. 36376 /note="assembly_fragment"
misc_feature	36477. 40247 /note="assembly_fragment"
misc_feature	40348. 82256 /note="assembly_fragment"
misc_feature	82357. 92140 /note="assembly_fragment"
misc_feature	92241. 108763 /note="assembly_fragment"
misc_feature	108864. 141973 /note="assembly_fragment"
misc_feature	142074. 170270 /note="assembly_fragment"
misc_feature	170371. 179269 /note="assembly_fragment"

Query Match 7.1%; Score 182.6; DB 2; Length 179269;  
 Best Local Similarity 82.6%; Pred. No. 3.2e-38;  
 Matches 209; Conservative 0; Mismatches 44; Indels 0; Gaps 0;  
 QY 687 AGGGAATATGAGGTCGAGGAGGACTGGAATGGTGGCCGCTTTTATTCCTCCCATCT 746  
 Db 46945 AGGGAATATGAGGTCGAGGAGGACTGGAATGGTGGCCGCTTTTATTCCTCCCATCT 47004  
 QY 747 TAATAATAAGTCTCGAGAGTGACATCTCTGTGTACGTGAAGATGGTCAGAGATTTCT 806  
 Db 47005 TAGTACAAATCATGCGAGAGTGACATCTCTGTGTACGTGAAGATGGTCAGAGATTTCT 47064  
 QY 807 CGTTAGTACTCTTCAGATTACATATATCTTTTGGACCCGAAAGATGATACAGCAGAGA 866

Db 47065 TGTCAGTATCTTCAGACATACATCTATCTTTTTCAGCCCAAGATGATGACTGCACGAGA 47124  
 QY 867 ACTTAAACTCTCTTCGCGAGAGAGAGAGAGAGTGGACACACACAGTTTAAACGG 926  
 Db 47125 ACTTAAACTCTCTTCGCGAGAGAGAGAGAGTGGACACACACAGTTTAAACGG 926  
 QY 927 TTGAGACTTCGT 939  
 Db 47185 CTTTCTCTTCTAGT 47197

RESULT 25  
 AC127803/c  
 LOCUS  
 DEFINITION Rattus norvegicus clone CH230-16L12, \*\*\* SEQUENCING IN PROGRESS  
 AC127803  
 VERSION 1  
 KEYWORDS HTG; HTGS\_PHASE1  
 SOURCE Rattus norvegicus  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 100694)  
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,  
 Albrooks,S.L., Amarantunga,H.C., Are,J.R., Ayie,M., Banks,T.,  
 Barbarella,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,  
 Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,  
 Bunay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
 Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
 Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
 Gabis,A., Gao,J., Garcia,A., Garner,T., Hale,S., Hamilton,K.,  
 Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Gill,R.,  
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,  
 Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,  
 Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,  
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
 Kravicic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
 Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulsegad,H.,  
 Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,  
 Massey,E., Mathiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,  
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,  
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,  
 Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,  
 Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savary,G.,  
 Scherer,S., Scott,G., Shen,H., Shoshitari,N., Sisson,I.,  
 Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H.,  
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
 Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,  
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,  
 Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K.,  
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
 Weinstein,G. and Gibbs,R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 100694)  
 Worley,K.C.  
 Direct Submission  
 Submitted (19-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

## COMMENT

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GXWC  
 Center clone name: CH230-16L12  
 ----- Summary Statistics  
 Sequencing vector: Plasmid;  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 37710 bases at least Q40  
 Consensus quality: 40353 bases at least Q30  
 Consensus quality: 42652 bases at least Q20  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 50 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 1772: contig of 1772 bp in length  
 \* 1773 1872: gap of unknown length  
 \* 1873 3356: contig of 1484 bp in length  
 \* 3357 3456: gap of unknown length  
 \* 3457 4772: contig of 1316 bp in length  
 \* 4773 4872: gap of unknown length  
 \* 4873 6219: contig of 1347 bp in length  
 \* 6220 6319: gap of unknown length  
 \* 6320 7626: contig of 1307 bp in length  
 \* 7627 7726: gap of unknown length  
 \* 7727 8787: contig of 1061 bp in length  
 \* 8788 9962: gap of unknown length  
 \* 9963 10062: contig of 1075 bp in length  
 \* 10063 11362: gap of unknown length  
 \* 11363 11462: contig of 1300 bp in length  
 \* 11463 12825: contig of 1362 bp in length  
 \* 12826 12924: gap of unknown length  
 \* 12925 14286: contig of 1362 bp in length  
 \* 14287 14386: gap of unknown length  
 \* 14387 15370: contig of 1183 bp in length  
 \* 15370 15669: gap of unknown length  
 \* 15670 16828: contig of 1159 bp in length  
 \* 16829 16928: gap of unknown length  
 \* 16929 18383: contig of 1455 bp in length  
 \* 18384 20334: gap of unknown length  
 \* 20335 20434: contig of 1851 bp in length  
 \* 20435 21570: contig of 1135 bp in length  
 \* 21570 21669: gap of unknown length  
 \* 21670 23401: contig of 1731 bp in length  
 \* 23401 23501: gap of unknown length  
 \* 23501 25107: contig of 1607 bp in length  
 \* 25108 25207: gap of unknown length  
 \* 25208 26488: contig of 1281 bp in length  
 \* 26489 26589: gap of unknown length  
 \* 26590 28082: contig of 1494 bp in length  
 \* 28083 28182: gap of unknown length  
 \* 28183 29847: contig of 1665 bp in length  
 \* 29848 31927: gap of unknown length  
 \* 31928 32027: contig of 1980 bp in length  
 \* 32028 33704: gap of unknown length  
 \* 33705 33804: contig of 1677 bp in length  
 \* 33805 35309: gap of unknown length  
 \* 35310 35409: contig of 1505 bp in length  
 \* 35410 36919: contig of 1510 bp in length

[illegible]





*	5738	7170: contig of 1433 bp in length
*	7171	7270: gap of 100 bp
*	7271	8657: contig of 1367 bp in length
*	8658	8757: gap of 100 bp
*	8758	10043: contig of 1286 bp in length
*	10044	10143: gap of 100 bp
*	10144	11873: contig of 1730 bp in length
*	11874	11973: gap of 100 bp
*	11974	13309: contig of 1336 bp in length
*	13310	13409: gap of 100 bp
*	13410	14723: contig of 1314 bp in length
*	14724	14823: gap of 100 bp
*	14824	16735: contig of 1312 bp in length
*	16736	16835: gap of 100 bp
*	16837	19688: contig of 2853 bp in length
*	19689	19788: gap of 100 bp
*	19789	22930: contig of 3142 bp in length
*	22931	23030: gap of 100 bp
*	23031	25546: contig of 2316 bp in length
*	25547	25646: gap of 100 bp
*	25647	28158: contig of 2512 bp in length
*	28159	28258: gap of 100 bp
*	28259	32365: contig of 4107 bp in length
*	32366	32465: gap of 100 bp
*	32466	36376: contig of 3911 bp in length
*	36377	36476: gap of 100 bp
*	36477	40247: contig of 3771 bp in length
*	40248	40347: gap of 100 bp
*	40348	82256: contig of 41909 bp in length
*	82257	82356: gap of 100 bp
*	82357	92140: contig of 9784 bp in length
*	92141	92240: gap of 100 bp
*	92241	108763: contig of 16523 bp in length
*	108764	108863: gap of 100 bp
*	108864	141973: contig of 33110 bp in length
*	141974	142073: gap of 100 bp
*	142074	170270: contig of 28197 bp in length
*	170271	170370: gap of 100 bp
*	170371	179269: contig of 889 bp in length.

## FEATURES

source

```
misc feature 1 773
/clone_lib="RPCI-23 Female Mouse BAC"
LOCUS
AC127803
```

	DEFINITION
misc feature	972 0724
	/note="assembly_fragment"

misc features	/note="assembly_fragment"	ACCESSION
2025	2025	VERSION
4010	4010	

	KEYWORDS	SOURCE
misc feature	/note="assembly_fragment"	
4211		
5022		

ORGANISM	note="assembly_fragment"
misq feature	
7700	
7700	

```

also feature=
707 355
/note="assembly_fragment"

```

```
/note="assembly_fragment"
```

```
/note="assembly_fragment"
```

```
/note="assembly_fragment"
```

```

/note="assembly_fragment"

```

```
20410: .14/23
/note="assembly_fragment"
```

```

1024. 10/33
/note="assembly fragment"

```

```
10030: ;19088
/note="assembly fragment"
```

```

"msb__sequence
1989: .22930
/note="assembly fragment"

```

```

"isc_feature
23031: .25346
/note="assembly fragment#

```

1

Query Match 6.2%; Score 159.6; DB 2; Length 179269;  
Best Local Similarity 90.0%;  
Pred. No. 6.1e-32;  
Matches 171; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

RESULT	28
AC127803	
LOCUS	AC127803
DEFINITION	Rattus norvegicus clone CH230-16L12, *** SEQUENCING IN PROGRESS ***, 50 unordered pieces.
	100694 bp DNA linear HTG 19-JUL-2002

AC127803  
AC127803.1 GI:21908185  
HTG: HTGS\_PHASE1.  
Rattus norvegicus.  
Rattus norvegicus.  
Rattus norvegicus.  
Eutamias; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1. (bases 1 to 100694)  
Muzny D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,  
Albrooks, J., Amaratunge, H.C., Are, J.R., Ayele, M., Banks, T.,  
Barbaria, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D.,  
Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,  
Buahy, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,  
Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,  
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,  
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,  
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,  
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
Ehrnath, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,  
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantcz, P.,  
Gabis, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,  
Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,

Homsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L. F.,  
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,  
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,  
 Kratovic, J., Kureishi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L.,  
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louised, H.,  
 Lozano, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,  
 Massey, E., Mahoney, E., McLeod, M. P., Meador, M., Mei, G., Metzger, M.,  
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,  
 Moser, M., Neal, D., Newton, S., Newton, N., Nguyen, A., Nguyen, N.,  
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G.,  
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
 Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y.,  
 Rivers, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savary, G.,  
 Scherer, S., Scott, G., Shen, H., Shoshitani, N., Sisson, I.,  
 Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,  
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
 Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,  
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
 Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,  
 Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstein, G., and Gibbs, R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 100694)  
 Worley, K. C.  
 Direct Submission  
 Submitted (19-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 -----  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 -----  
 Project Information  
 Center project name: GXYC  
 Center clone name: CH230-16L12  
 -----  
 Summary Statistics  
 Sequencing vector: Plasmid  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 37710 bases at least Q40  
 Consensus quality: 40353 bases at least Q30  
 Consensus quality: 42652 bases at least Q20  
 -----  
 NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 50 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \*  
 1 1772: contig of 1772 bp in length  
 1773: gap of unknown length  
 1873: contig of 1484 bp in length  
 3356: gap of unknown length  
 3456: contig of 1316 bp in length  
 4772: gap of unknown length  
 4773: contig of 1347 bp in length  
 4873: gap of unknown length  
 6220: contig of 1307 bp in length  
 6320: gap of unknown length  
 7626: contig of 1061 bp in length  
 7727: gap of unknown length  
 8887: contig of 1075 bp in length  
 8888: gap of unknown length  
 9963: contig of 1040 bp in length  
 10063: contig of 1300 bp in length  
 11362: gap of unknown length  
 11363: gap of unknown length  
 11462: contig of 2830 bp in length  
 11463: contig of 1362 bp in length  
 12824: gap of unknown length  
 12825: contig of 1362 bp in length  
 14286: gap of unknown length  
 14386: gap of unknown length  
 14387: contig of 1183 bp in length  
 15569: gap of unknown length  
 15570: contig of 1159 bp in length  
 16828: gap of unknown length  
 16829: contig of 1455 bp in length  
 16929: gap of unknown length  
 18384: contig of 1851 bp in length  
 20334: gap of unknown length  
 20335: contig of 1135 bp in length  
 21569: gap of unknown length  
 21570: contig of 1731 bp in length  
 23400: gap of unknown length  
 23500: contig of 1607 bp in length  
 25107: gap of unknown length  
 25207: contig of 1281 bp in length  
 26488: gap of unknown length  
 26589: contig of 1494 bp in length  
 28182: gap of unknown length  
 28183: contig of 1665 bp in length  
 29847: gap of unknown length  
 29848: contig of 1980 bp in length  
 31927: gap of unknown length  
 32027: contig of 1677 bp in length  
 33704: gap of unknown length  
 33804: contig of 1505 bp in length  
 35309: gap of unknown length  
 35409: contig of 1510 bp in length  
 36919: gap of unknown length  
 37019: contig of 1380 bp in length  
 38399: gap of unknown length  
 38499: contig of 1184 bp in length  
 39683: gap of unknown length  
 39783: contig of 1047 bp in length  
 40830: gap of unknown length  
 40930: contig of 1881 bp in length  
 42811: gap of unknown length  
 42911: contig of 1109 bp in length  
 44020: gap of unknown length  
 44120: contig of 1637 bp in length  
 45757: gap of unknown length  
 45857: contig of 1794 bp in length  
 47651: gap of unknown length  
 47751: contig of 1289 bp in length  
 49040: gap of unknown length  
 49140: contig of 2234 bp in length  
 51374: gap of unknown length  
 51474: contig of 2178 bp in length  
 53652: gap of unknown length  
 53752: contig of 1690 bp in length  
 55442: gap of unknown length  
 55443: contig of 2423 bp in length  
 57965: gap of unknown length  
 58065: contig of 1596 bp in length  
 59661: gap of unknown length  
 59761: contig of 2261 bp in length  
 62022: gap of unknown length  
 62122: contig of 2591 bp in length  
 64713: gap of unknown length  
 64813: contig of 3105 bp in length  
 67918: gap of unknown length  
 68018: contig of 3105 bp in length  
 71123: gap of unknown length  
 71223: contig of 2707 bp in length  
 73930: gap of unknown length  
 74030: contig of 1974 bp in length  
 76004: gap of unknown length  
 76104: contig of 1040 bp in length  
 77144: gap of unknown length  
 77244: contig of 2830 bp in length  
 80074: contig of 2830 bp in length

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT



```

* 80075 80174: gap of unknown length
* 80175 83578: contig of 3404 bp in length
* 83579 83579: gap of unknown length
* 83679 86140: contig of 2462 bp in length
* 86141 86240: gap of unknown length
* 86241 89933: contig of 3693 bp in length
* 89934 90034: gap of unknown length
* 90034 94840: contig of 4807 bp in length
* 94841 94940: gap of unknown length
* 94941 100694: contig of 5754 bp in length.
      Location/Qualifiers
        1. 100694
          /organism="Rattus norvegicus"
          /db_xref="taxon:10116"
          /clone="CH230-16L12"
BASE COUNT 25190 a 18744 c 18830 g 24955 t 12975 others
ORIGIN

Query Match          5.5%; Score 142.4; DB 2; Length 100694;
Best Local Similarity 93.1%; Pred. No. 2.9e-27;
Matches 149; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 2166 GATAAAGAGCCAAATTCCTGGGGTCTAACTTTGTAATGAGTGGTTCTGACTGTGGCCA 2225
|||||
Db 76852 GATAAAGAGCCAAATTCCTGGGGTCTAACTTTGTAATGAGTGGTTCTGACTGTGGCCA 76911
|||||

QY 2226 CATTTTCATCTGGGATCGGCACACTGCTGAGCATTTGATGCTTCTGGAAGCTGATAATCA 2285
|||||
Db 76912 TATCTTCATCTGGGACCGGCACACTGCTGAGCATTTGATGCTTCTGGAAGCTGATAATCA 76971
|||||

QY 2286 TCTGTTAACTGCTGCGACCCACATCCGTTTGACCAAT 2325
|||||
Db 76972 CCGTGTCAACTGCTGCGACCCACATCCATTCATTCGACCAAGT 77011
|||||

RESULT 29
AC100993/c
LOCUS          AC100993          59143 bp    DNA          linear    HTG 23-NOV-2001
DEFINITION    Mus musculus clone RP23-77G13, LOW-PASS SEQUENCE SAMPLING.
ACCESSION     AC100993
VERSION       AC100993.1 GI:17059767
KEYWORDS      HTG; HTGS_PHASED.
SOURCE        Mus musculus
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 59143)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Mus musculus, clone RP23-77G13
Unpublished
2 (bases 1 to 59143)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Choepe, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
Ferrelira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-pierre, N.,
Hagos, B., Hearford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K.,
Lamazares, K., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPeckers, R., Meldrim, J.,
Meneus, L., Mihova, T., Mlenka, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Telamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Triglio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, N.

```

TITLE  
JOURNAL

## COMMENT

Direct Submission  
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
-----  
Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
-----  
Project Information  
Center project name: L14696  
Center clone name: 77\_G\_13  
-----

\* NOTE: This record contains 74 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 712: contig of 712 bp in length  
813 812: gap of 100 bp  
813 1502: contig of 690 bp in length  
1503 1602: gap of 100 bp  
1603 2325: contig of 723 bp in length  
2326 2425: gap of 100 bp  
2426 3145: contig of 720 bp in length  
3146 3245: gap of 100 bp  
3246 3906: contig of 661 bp in length  
3907 4006: gap of 100 bp  
4007 4708: contig of 702 bp in length  
4709 4808: gap of 100 bp  
4809 5483: contig of 675 bp in length  
5484 5583: gap of 100 bp  
5584 6277: contig of 694 bp in length  
6278 6377: gap of 100 bp  
6378 7081: contig of 704 bp in length  
7082 7181: gap of 100 bp  
7182 7899: contig of 718 bp in length  
7900 7999: gap of 100 bp  
8000 8689: contig of 690 bp in length  
8690 8789: gap of 100 bp  
8790 9517: contig of 728 bp in length  
9518 9617: gap of 100 bp  
9618 10249: contig of 632 bp in length  
10250 10349: gap of 100 bp  
10350 11057: contig of 708 bp in length  
11058 11157: gap of 100 bp  
11158 11850: contig of 693 bp in length  
11851 11950: gap of 100 bp  
11951 12621: contig of 671 bp in length  
12622 12721: gap of 100 bp  
12722 13417: contig of 696 bp in length  
13418 13517: gap of 100 bp  
13518 14208: contig of 691 bp in length  
14209 14308: gap of 100 bp  
14309 15001: contig of 693 bp in length  
15002 15101: gap of 100 bp  
15102 15795: contig of 694 bp in length  
15796 15895: gap of 100 bp  
15896 16595: contig of 700 bp in length  
16596 16695: gap of 100 bp  
16696 17432: contig of 737 bp in length  
17433 17532: gap of 100 bp  
17533 18233: contig of 701 bp in length  
18234 18333: gap of 100 bp  
18334 19056: contig of 723 bp in length

```
* 19057 19156: gap of 100 bp
* 19157 19852: contig of 896 bp in length
* 19853 19952: gap of 100 bp
* 19953 20667: contig of 715 bp in length
* 20668 20767: gap of 100 bp
* 20768 21449: contig of 682 bp in length
* 21450 21549: gap of 100 bp
* 21550 22161: contig of 612 bp in length
* 22162 22261: gap of 100 bp
* 22262 22957: contig of 696 bp in length
* 22958 23057: gap of 100 bp
* 23058 23779: contig of 722 bp in length
* 23780 23879: gap of 100 bp
* 23880 24591: contig of 712 bp in length
* 24592 24691: gap of 100 bp
* 24692 25418: contig of 727 bp in length
* 25419 25518: gap of 100 bp
* 25519 26226: contig of 708 bp in length
* 26227 26326: gap of 100 bp
* 26327 27019: contig of 693 bp in length
* 27020 27119: gap of 100 bp
* 27120 27814: contig of 695 bp in length
* 27815 27914: gap of 100 bp
* 27915 28612: contig of 698 bp in length
* 28613 28712: gap of 100 bp
* 28713 29436: contig of 724 bp in length
* 29437 29536: gap of 100 bp
* 29537 30238: contig of 702 bp in length
* 30239 30338: gap of 100 bp
* 30339 31043: contig of 705 bp in length
* 31044 31143: gap of 100 bp
* 31144 31881: contig of 738 bp in length
* 31882 31981: gap of 100 bp
* 31982 32694: contig of 713 bp in length
* 32695 32794: gap of 100 bp
* 32795 33478: contig of 684 bp in length
* 33479 33578: gap of 100 bp
* 33579 34276: contig of 698 bp in length
* 34277 34376: gap of 100 bp
* 34377 35074: contig of 698 bp in length
* 35075 35174: gap of 100 bp
* 35175 35878: contig of 704 bp in length
* 35879 35978: gap of 100 bp
* 35979 36661: contig of 683 bp in length
* 36662 36761: gap of 100 bp
* 36762 37561: contig of 700 bp in length
* 37562 38274: contig of 713 bp in length
* 38275 38374: gap of 100 bp
* 38375 39052: contig of 678 bp in length
* 39053 39152: gap of 100 bp
* 39153 39852: contig of 700 bp in length
* 39853 39952: gap of 100 bp
* 39953 40654: contig of 702 bp in length
* 40655 40754: gap of 100 bp
* 40755 41360: contig of 606 bp in length
* 41361 41460: gap of 100 bp
* 41461 42155: contig of 695 bp in length
* 42156 42255: gap of 100 bp
* 42256 42953: contig of 698 bp in length
* 42954 43053: gap of 100 bp
* 43054 43774: contig of 721 bp in length
* 43775 43874: gap of 100 bp
* 43875 44596: contig of 722 bp in length
* 44597 44696: gap of 100 bp
* 44697 45410: contig of 714 bp in length
* 45411 45510: gap of 100 bp
* 45511 46201: contig of 691 bp in length
* 46202 46301: gap of 100 bp
* 46302 47012: contig of 711 bp in length
* 47013 47112: gap of 100 bp
* 47113 47818: contig of 706 bp in length
* 47819 47918: gap of 100 bp
*
* 47919 48611: contig of 693 bp in length
* 48612 48711: gap of 100 bp
* 48712 49415: contig of 704 bp in length
* 49416 49515: gap of 100 bp
* 49516 50231: contig of 716 bp in length
* 50232 50331: gap of 100 bp
* 50332 51060: contig of 729 bp in length
* 51061 51160: gap of 100 bp
* 51161 51899: contig of 739 bp in length
* 51900 51999: gap of 100 bp
* 52000 52719: contig of 720 bp in length
* 52720 52819: gap of 100 bp
* 52820 53540: contig of 721 bp in length
* 53541 53640: gap of 100 bp
* 53641 54332: contig of 692 bp in length
* 54333 54432: gap of 100 bp
* 54433 55139: contig of 707 bp in length
* 55140 55239: gap of 100 bp
*
Query Match 4.9%; Score 125.2; DB 2; Length 59143;
Best Local Similarity 70.4%; Pred. No. 1.4e-22;
Matches 197; Conservative 0; Mismatches 78; Indels 5; Gaps 2;
QY 1 ATGCTCGGGTGGCTCTACCCACACCTGTTGGGACGTGAGGAAAGTCCCTCGGG 60
DB 1138 AAGTTTATAGTGTATTTCTGCTGCACCTGTATAGGACATGAGGAAAGAACCTCAGG 1079
QY 61 CTGAGGAGCCCTCCCGGCTGCGAGTGCCTACCTGGGAGAGAGATTTATCCAAAGA 120
DB 1078 CTCAAAAGTCTCTCCCTCTGAGAGCCACTGCCTG---AGAAGAGACTTAGGCAAAACA 1022
QY 121 TTAACACTTGAAGCAACCTTAATGTGCATGATGGTGTGTTTACATCAATCTGTTGGAAT 180
DB 1021 CTAAGCCTTCAGGCACATCAATGTTCATAAGCAGGTTAGTAGT--GATCTTTTAGAAG 964
QY 181 GACACTGGAGATATATTTATCTGGCTCAGATGACACCAAAATTAATTAATTAATCT 240
DB 963 GACACTAGCAATATATATTTTGTGCTTAGATGATGCGCAAGTTAGTAGTTAGTATGCT 904
QY 241 TACAGCAGAAAGTTTGGACAAATTCGTTTCAGGCAACC 280
DB 903 CACAGCAGAAAGGCTTTGACACAAATCATCCAAAGGCACC 864
*
RESULT 30
AL732543 232993 bp DNA linear HTG 17-AUG-2002
LOCUS Mus musculus chromosome 4 clone RP23-13D22, *** SEQUENCING IN
DEFINITION PROGRESS ***, 5 unordered pieces.
ACCESSION AL732543
VERSION AL732543.2 GI:22415946
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 232993)
AUTHORS Direct Submission
TITLE Submitted (16-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
JOURNAL Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Aug 21, 2002 this sequence version replaced gi:21213781.
COMMENT ----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: bML3D22
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
```

Consensus quality: 231947 bases at least Q40  
 Consensus quality: 232218 bases at least Q30  
 Consensus quality: 232424 bases at least Q20  
 Insert size: 232593; sum-of-contigs  
 Insert size: 233521; 6.4% error; agarose-fp  
 Quality coverage: 5.47x in Q20 bases; sum-of-contigs Quality  
 coverage: 5.45x in Q20 bases; agarose-fp

NOTE: This is a 'working draft' sequence. It currently  
 consists of 5 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.

1 131776: contig of 131776 bp in length  
 131777 131876: gap of 100 bp  
 131877 180768: contig of 48892 bp in length  
 180769 180868: gap of 100 bp  
 180869 227965: contig of 47097 bp in length  
 227966 228065: gap of 100 bp  
 228066 230501: contig of 2436 bp in length  
 230502 230601: gap of 100 bp  
 230602 232993: contig of 2392 bp in length.

## FEATURES

source

1. 232993  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /chromosome="4"  
 /clone="RP23-13D22"  
 /clone\_lib="RPCI-23"

misc\_feature

1. 131776  
 /note="assembly\_fragment:01399  
 fragment\_chain:1  
 clone\_end:SP6  
 vector\_side:left"

misc\_feature

131877..180768  
 /note="assembly\_fragment:01212  
 fragment\_chain:1"

misc\_feature

180869..227965  
 /note="assembly\_fragment:00644  
 fragment\_chain:1"

misc\_feature

228066..230501  
 /note="assembly\_fragment:02215  
 fragment\_chain:2"

misc\_feature

230602..232993  
 /note="assembly\_fragment:02207  
 fragment\_chain:2"

BASE COUNT

67150 a 50371 c 47633 g 67439 t 400 others

ORIGIN

Query Match 4.9%; Score 125.2; DB 2; Length 232993;  
 Best Local Similarity 70.4%; Pred. No. 1.6e-22;  
 Matches 197; Conservative 0; Mismatches 78; Indels 5; Gaps 2;  
 QY 1 ATGTCGGGGTGGCTCTACCCACACCTGTTGTGGGACGTGAGGAAAGTCCCTCGG 60  
 Db 46014 AAGTTTATAGTGAATTTCTGCTGCACCTGTTATAGGACATGAGGAAAGACCTCAGG 46073  
 QY 61 CTGGAGGACCGTCGCCGCTCGGAGTCCCTACCTGGGAGGAGAAATTTATCCAAAGA 120  
 Db 46074 CTACAAAGCTGTCCTCCTGAAGAGCCACTGCCTG---AGAAGAGACTTAGGCAAAACA 46130  
 QY 121 TTAACACTTGAAGCAACCTTTAATGTGTCATGATGTTGTGTTTATATACATCTGTTGGAT 180  
 Db 46131 CTAAGCCTTCAGGCAACATGAATGTTATCAAGCAGGTAGTAGT--GATCTTTTAGAAG 46188  
 QY 181 GACACTGGAGATATATTTATCTGGCTCAGATGACACCAATATAGTAATAGTAATCCT 240  
 Db 46189 GACACTGACAAATATATTTCTGCTGGCTAGATGATGCAAGTAGTAGTAGTAGTCT 46248  
 QY 241 TACAGCAGAAAGGTTTGTGACAACAATTCGTTTCAGGGCACC 280

Db 46249 CACAGCAGAAAGCTTTGACACAAATCATCCAGGCACC 46288  
 RESULT 31  
 AC069147/c

LOCUS

DEFINITION

AC069147 158324 bp DNA linear HTG 01-SEP-2000  
 Homo sapiens chromosome 1 clone RP11-105H9, WORKING DRAFT SEQUENCE,  
 7 unordered pieces.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AC069147 158324 bp DNA linear HTG 01-SEP-2000  
 Homo sapiens chromosome 1 clone RP11-105H9, WORKING DRAFT SEQUENCE,  
 7 unordered pieces.  
 AC069147 3 GI:8099949  
 HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 Homo sapiens.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 Waterston,R.H. 1 (bases 1 to 158324)  
 The sequence of Homo sapiens clone  
 Unpublished  
 2 (bases 1 to 158324)  
 Waterston,R.H.  
 Direct Submission  
 Submitted (19-MAY-2000) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 On May 28, 2000 this sequence version replaced gi:8028173.

----- Genome Center -----  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: http://genome.wustl.edu/gsc/index.shtml

----- Project Information -----  
 Center project name: H\_NH0105H09

----- Summary Statistics -----  
 Sequencing vector: M13; 100%  
 Chemistry: Dye-primer ET; 100% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 154865 bases at least Q40  
 Consensus quality: 155623 bases at least Q30  
 Consensus quality: 156154 bases at least Q20  
 Insert size: 174000; agarose-fp  
 Insert size: 157724; sum-of-contigs  
 Quality coverage: 7.55 in Q20 bases; agarose-fp  
 Quality coverage: 8.39 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently  
 consists of 7 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.

1 6587: contig of 6587 bp in length  
 6588 6687: gap of unknown length  
 6688 14247: contig of 7560 bp in length  
 14248 14347: gap of unknown length  
 14348 28850: contig of 14503 bp in length  
 28851 28950: gap of unknown length  
 28951 49498: contig of 20548 bp in length  
 49499 49598: gap of unknown length  
 49599 69546: contig of 19948 bp in length  
 69547 105850: contig of 36204 bp in length  
 105851 105950: gap of unknown length  
 105951 158324: contig of 52374 bp in length.

Location/Qualifiers  
 1. 158324  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"

FEATURES  
 source

```

/misc_feature 1. 6587 /chromosome="1"
/misc_feature 6688. 14247 /clone="RP11-105H9"
/misc_feature 14348. 28850 /note="assembly_name:Contig15"
/misc_feature 14348. 28850 /note="assembly_name:Contig16"
/misc_feature 14348. 28850 /note="assembly_name:Contig17"
/misc_feature 28951. 49498 /note="assembly_name:Contig18"
/misc_feature 49599. 69546 /note="assembly_name:Contig19"
/misc_feature 69647. 105850 /note="assembly_name:Contig20"
/misc_feature 105951. 158324 /note="assembly_name:Contig21"
BASE COUNT 47660 a 30780 c 31018 g 48209 t 657 others
ORIGIN
Query Match 3.8%; Score 98; DB 2; Length 158324;
Best Local Similarity 100.0%; Pred. No. 4e-15;
Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGTCGGGGTGGCTCCTACCCACACTGTGTGGGACGTGAGGAAAGTCCCTCGGG 60
Db 93904 ATGTCGGGGTGGCTCCTACCCACACTGTGTGGGACGTGAGGAAAGTCCCTCGGG 93845
QY 61 CTGGAGACCCGTCCTCCGCTCGGAGTGCCTACCTGGG 98
Db 93844 CTGGAGACCCGTCCTCCGCTCGGAGTGCCTACCTGGG 93807

RESULT 32
HS295C6 102258 bp DNA linear PRI 23-NOV-1999
LOCUS Human DNA sequence from PAC 295C6 on chromosome 1q24. Contains
DEFINITION ESTs, CA repeat, STS and CpG island.
ACCESSION 297876
VERSION 297876.1 GI:2582745
KEYWORDS 1q24; CpG island; repeat polymorphism.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Graham, D.
Direct Submission
Submitted (27-OCT-1997) Chromosome 1 Project Group
(http://www.sanger.ac.uk/HGP/Chr1/) Sanger Centre, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquires:
humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Nov 2, 1997 this sequence version replaced gi:2465042.
IMPORTANT: This sequence is the entire insert of clone 295C6.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variations annotated may not be found in the sequence submission
corresponding to the overlapping clone as we submit sequences with
only a small overlap as described above.
This sequence was generated from part of bacterial clone contigs of
human chromosome 1, constructed by the Sanger Centre chromosome 1
mapping group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1/
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.

```

The true left end of clone 295C6 is at 1 in this sequence. The true right end of clone 295C6 is at 102258.  
295C6 is from the library RPI1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong.  
For further details see <http://bacpac.med.buffalo.edu/>.

## FEATURES

## source

```

Location/Qualifiers
1. 102258
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="1q24"
/clone="RP1-295C6"
/clone_lib="RPI-1"
891. 1180
/note="AluSq repeat: matches 298. 9 of consensus"
1375. 1683
/note="AluSp repeat: matches 303. 1 of consensus"
2028. 2310
/note="AluSq repeat: matches 300. 1 of consensus"
2612. 2903
/note="AluSx repeat: matches 301. 1 of consensus"
complement(4475..4866)
/note="match: Z23618 STS containing (CA) repeat"
4625. 4670
/note="23 copies of GT 100 % conserved; differs from Z23618"
4766. 4897
/note="MIR2 repeat: matches 145. 1 of consensus"
4969. 5140
/note="MER42c repeat: matches 1475. 1300 of consensus"
5146. 5448
/note="AluJo repeat: matches 302. 1 of consensus"
5303. 5795
/note="AluSg repeat: matches 1. 299 of consensus"
5797. 5925
/note="MER42c repeat: matches 1265. 1124 of consensus"
5923. 6039
/note="L1MB6 repeat: matches 920. 804 of consensus"
6123. 6415
/note="AluSx repeat: matches 1. 292 of consensus"
7799. 8099
/note="AluSx repeat: matches 302. 1 of consensus"
9002. 9302
/note="AluSp repeat: matches 2. 303 of consensus"
9692. 9983
/note="AluJo repeat: matches 1. 301 of consensus"
10986. 11071
/note="MIR2 repeat: matches 146. 60 of consensus"
11583. 11766
/note="MIR repeat: matches 262. 70 of consensus"
12180. 12479
/note="AluSg repeat: matches 1. 300 of consensus"
13246. 13500
/note="MIR repeat: matches 256. 1 of consensus"
13907. 13973
/note="MIR repeat: matches 75. 141 of consensus"
13975. 14489
/note="MER1A repeat: matches 526. 1 of consensus"
15702. 16064
/note="PHELC repeat: matches 1. 371 of consensus"
16144. 16585
/note="MIR1C repeat: matches 9. 466 of consensus"
17173. 17534
/note="LNR2 repeat: matches 105. 449 of consensus"
<17786..>18427
/note="match: multiple ESTs; match: AA165668 C16515
AA077391"
18004. 18120
/note="3 copies of 39 mer 81 % conserved"
19377. 21532
/note="match: multiple ESTs; match: N46026 H60052 H12822
T62974 AA283144; match: T57835 F00049 N75628 AA382351
AA300207; match: W37181 N52820 AA337499 W87891 H84729;

```

repeat\_region

repeat\_region

repeat\_region

repeat\_region

misc\_feature

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

prim\_transcript

repeat\_region

prim\_transcript

similar to endogenous retrovirus POL POLYPROTEIN"  
 <22156..22928  
 /note="match: multiple ESTs; match: AA401243 AA258918  
 AA248892 AA096209 R36280"  
 repeat\_region 23025..23060  
 /note="3 copies of 12 mer 94 % conserved"  
 repeat\_region 23519..23865  
 /note="LTR2 repeat: matches 105..449 of consensus"  
 repeat\_region 24473..24785  
 /note="AluSq repeat: matches 1..299 of consensus"  
 repeat\_region 25037..25058  
 /note="16 copies of 2 mer 88 % conserved"  
 repeat\_region 25877..25987  
 /note="L1P45 repeat: matches 788..892 of consensus"  
 repeat\_region 26032..26312  
 /note="AluSp repeat: matches 1..295 of consensus"  
 repeat\_region 26752..27050  
 /note="AluSq repeat: matches 1..299 of consensus"  
 repeat\_region 28012..28312  
 /note="AluSq repeat: matches 1..303 of consensus"  
 repeat\_region 28316..29066  
 /note="L1ME3A repeat: matches 109..907 of consensus"  
 repeat\_region 29517..29629  
 /note="MIR repeat: matches 262..151 of consensus"  
 repeat\_region 29631..29928  
 /note="AluY repeat: matches 1..295 of consensus"  
 repeat\_region 29937..30237  
 /note="AluSq repeat: matches 1..301 of consensus"  
 repeat\_region 30238..30315  
 /note="MIR repeat: matches 158..76 of consensus"  
 repeat\_region 31096..31287  
 /note="MIR repeat: matches 56..262 of consensus"  
 repeat\_region 31289..31379  
 /note="MIR2 repeat: matches 146..56 of consensus"  
 repeat\_region 31479..31525  
 /note="MIR2 repeat: matches 146..95 of consensus"  
 repeat\_region 33069..33318  
 /note="MIR repeat: matches 9..253 of consensus"  
 repeat\_region 35365..35744  
 /note="MIR repeat: matches 262..74 of consensus"  
 repeat\_region 36165..36466  
 /note="AluSq repeat: matches 1..302 of consensus"  
 repeat\_region 36692..36944  
 /note="MIR repeat: matches 3..240 of consensus"  
 repeat\_region 36999..37299  
 /note="AluSq repeat: matches 301..1 of consensus"  
 repeat\_region 37924..38114  
 /note="L1MA2 repeat: matches 1055..866 of consensus"  
 repeat\_region 38108..38475  
 /note="MSTP repeat: matches 426..1 of consensus"  
 repeat\_region 38479..38657  
 /note="L1MA2 repeat: matches 874..687 of consensus"  
 repeat\_region 38658..38957  
 /note="AluSq repeat: matches 300..1 of consensus"  
 repeat\_region 38958..39580  
 /note="L1MA3 repeat: matches 699..85 of consensus"  
 repeat\_region 39585..39875  
 /note="AluSq repeat: matches 292..1 of consensus"  
 repeat\_region 39877..39951  
 /note="L1MA9 repeat: matches 88..14 of consensus"  
 repeat\_region 39944..40919  
 /note="L1 repeat: matches 4416..5390 of consensus"  
 repeat\_region 40769..41662  
 /note="L1P42 repeat: matches 1..893 of consensus"  
 repeat\_region 41764..42556  
 /note="L1 repeat: matches 5133..4305 of consensus"  
 repeat\_region 42557..42857  
 /note="AluY repeat: matches 1..301 of consensus"  
 repeat\_region 42876..43925  
 /note="L1 repeat: matches 4313..3215 of consensus"  
 repeat\_region 44235..44498  
 /note="MER43 repeat: matches 3..272 of consensus"  
 44788..44860

/note="MIR repeat: matches 154..82 of consensus"  
 44933..45220  
 /note="AluSq repeat: matches 289..2 of consensus"  
 45304..45336  
 repeat\_region 45337..45637  
 /note="AluY repeat: matches 1..301 of consensus"  
 45934..46220  
 /note="AluYb repeat: matches 302..14 of consensus"  
 46817..46935  
 /note="MIR repeat: matches 35..154 of consensus"  
 47393..47677  
 /note="AluSq repeat: matches 1..302 of consensus"  
 48309..48404  
 /note="MIR repeat: matches 48..140 of consensus"  
 48620..48862  
 /note="MER21B repeat: matches 347..102 of consensus"  
 48865..49159  
 /note="AluSc repeat: matches 1..296 of consensus"  
 51182..51467  
 Query Match 3.7%; Score 96.4; DB 9; Length 102258;  
 Best Local Similarity 99.0%; Pred. No. 1.1e-14;  
 Matches 97; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 ATGTCTCGGGTGGCTCTACCCACACCTGTGTGGACGTGAGGAAAGGTCCTCGGG 60  
 |||||  
 Db 90874 ATGTCTCGGGTGGCTCTACCCACACCTGTGTGGACGTGAGGAAAGGTCCTCGGG 90933  
 |||||  
 QY 61 CTGGAGGACCGTCCCGGCTCGGAGTGCTGACCTGGG 98  
 |||||  
 Db 90934 CTGGAGGACCGTCCCGGCTCGGAGTGCTGACCTGGG 90971  
 |||||

RESULT 33  
 AY120722

LOCUS  
 DEFINITION

AY120722 1765 bp mRNA linear PLN 22-JUN-2002  
 Arabidopsis thaliana putative protein (At4g35140) mRNA, complete cds.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

AY120722  
 AY120722.1 GI:21539454  
 FLI\_CDNA.  
 thale cress.

Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

REFERENCE  
 AUTHORS

1 (bases 1 to 1765)  
 Southwick, A., Karlin-Neumann, G., Nguyen, M., Tripp, M., Miranda, M.,  
 Palm, C.J., Bowser, L., Jones, T., Banh, J., Carninci, P., Chen, H.,  
 Cheuk, R., Chung, M.K., Hayashizaki, Y., Ishida, J., Kamiya, A.,  
 Kawai, J., Kim, C., Lin, J., Liu, S.X., Narusaka, M., Pham, P.K.,  
 Sakano, H., Sakurai, T., Satou, M., Seki, M., Shinn, P., Yamada, K.,  
 Shinozaki, K., Ecker, J., Theologis, A. and Davis, R.W.  
 Direct Submission

TITLE  
 JOURNAL

Submitted (11-JUN-2002) DNA Sequencing and Technology Center,  
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,  
 USA

COMMENT

e-mail for correspondence: arab@sequence.stanford.edu

RIKEN Genomic Sciences Center (GSC) members carried out the  
 collection and clustering of RAFL cDNAs (RAFL CDNA : 'RIKEN  
 Arabidopsis Full-length cDNA'): Seki, M., Narusaka, M., Ishida, J.,  
 Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,  
 Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGEC (SSP) Consortium members carried out the  
 sequencing and annotation of the RAFL cDNAs: Southwick, A.,  
 Nguyen, M., Karlin-Neumann, G., Lam, B., Miranda, M., Palm, C.J.,  
 Bowser, L., Jones, T., Banh, J., Chen, H., Cheuk, R., Chung, M.K.,  
 Kim, C., Lin, J., Liu, S.X., Pham, P.K., Sakano, H., Shinn, P.,  
 Yamada, K., Ecker, J., Theologis, A. and Davis, R.W.

Southwick, A. (SSP/Stanford) and Seki, M. (RIKEN GSC) contributed









```

Db 1908 GGCACACTGCAACACACACCGGATATCAAGAGGCCAAATTTCTTTGGCAGCAACGCTCAG 1967
Qy 2197 TTTGTAATGAGTGGTCTGACACTGTGGCCACATTTTCATCTGGGATCGGCACACTGCTGAG 2256
Db 1968 TATATCGTCACTGGGCTCTGACGACTGCTCTCTTCATCTGGGAAAGAGAGACACCAAC 2027
Qy 2257 CATTTGATGCTTCTGGAAGCTGATATCATGTTGGTAACTGCTGCGAGCCACATCCGTTT 2316
Db 2028 CTGGTCCGCTGCTCTCAAGGAGGATGAGTCCATTTGCAACTGCTGCTGAGCCACACCCAGC 2087
Qy 2317 GACCCAAATTTAGCTCATCTGCGCATAGATTATGACATAAAGATCTGTCACCATTAGAA 2376
Db 2088 TACTGCTTCTGGCCACGACGATGCTGCTGTTGTCGGGCTCTGGAACCCCGACCA 2147
Qy 2377 GAG 2379
Db 2148 GAG 2150

RESULT 39
BC032523
LOCUS BC032523 4305 bp mRNA linear PRI 26-JUN-2002
DEFINITION Homo sapiens, clone MGC:44910 IMAGE:5500384, mRNA, complete cds.
ACCESSION BC032523
VERSION BC032523.1 GI:21595707
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Strausberg, R.
Direct Submission
Submitted (06-JUN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: gcgaps-remail.nih.gov
Tissue Procurement: Lou Staudt
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc.mgc@nih.gov
Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,
Maduro, Q.L., Masello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Young, A., Zhang, L.-H. and Green, E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 69 Row: 9 Column: 7
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 14149658.
Location/Qualifiers
1. 4305
/organism="Homo sapiens"
/db_xref="LocusID:230796"
/db_xref="taxon:9606"
/clone="MGC:44910 IMAGE:5500384"
/tissue_type="Lymph, lymphoma"
/clone_lib="NIH_MGC_85"
/lab_host="DH10B"
note="Vector: pCMV-SPORT6"
399. .2429

FEATURES
source
CDS

```

```

/codon_start=1
/product="Unknown (protein for MGC:44910)"
/protein_id="AAH32523.1"
/db_xref="GI:21595708"
/translation="MAKVNITRDLIRRIKQKALSPERRYVHTDPTFIRRLGLEAELQ
GHSGCVCNLEWNEKGDILASGSDOHTIWDPLHKKLLSMHTGHTANIFSVKFLPFA
QDRILLTGADSKVHVHDLTVKETHMFQDHTNRKVRATATAMPPTWFAAEDGLIR
OYDLRENSKSEVLIDLEYCGQLVEAKCLTVNPDNNCLAVGASGPFVRLYDIEMIH
NHRKSMKQSPAGVHTCDROKPLDGAAYVAGHLFVKLPDYNNRRLVAVTVTF
SPHTELLVNMGGQVYLFDTLYKRPYTFELLPRKSHSSQVQNGKMTNGSVNGSNG
LHLSNGFRLPESGHVSPQVPELPYLLERVKQOANEAFACOOQTQAIOLYKAVQORAP
HNALYGNRAAYMKRWKGDHVDALDCLKAI SLNPKHLKAHRLARCLPELKYVAE
ALCLDDEKGFKEQAHSSADALGRDITAFSLKNDGEEKKGGFGGAPVRLSRK
DSTSEDEMVLRESDYQFRYCHCNTTIDKEANFFGGSNAQYIVSGSDGSGFTWEK
ETNLVRYLQGDSEIVNCLQPHSYCFATSGIDPVVRLMNPRESDELTRGVVDEME
GASQANORMNADPLEVMLLNMGYRIITGLSSGGAGASDDESSGQVQCRPS"
BASE COUNT 858 a 1343 c 1189 g 915 t
ORIGIN
Query Match 2.7% Score 70.6; DB 9; Length 4305;
Best Local Similarity 58.88; Pred. No. 9e-08;
Matches 143; Conservative 0; Mismatches 94; Indels 6; Gaps 1;
QY 2143 GGCCATCGCAACTCCAGGACAAATGATAAAGAGCCAAATTTCTGGG-----TGCTAAC 2196
Db 1989 GGCCACTGCAACACACACCGGATATCAAGAGGCCAAATTTCTTTGGCAGCAACCTCAG 2048
QY 2197 TTTGTAATGAGTGGTCTGACTGTGGCCACATTTTCATCTGGGATCGGCACACTGCTGAG 2256
Db 2049 TATATCGTCACTGGCTGCTGACGATGGCTCTTCTTCATCTGGGAAAGAGGAGACCAAC 2108
QY 2257 CATTTGATGCTTCTGGAAGCTGATAATCATGTGTAACCTGCTGCAGCCACATCCGTTT 2316
Db 2109 CTGTCCTGCTGCTCCAGGGGATGATCCATTTGCTCACTGCTGCAGCCACACCCAGC 2168
QY 2317 GACCCAAATTTAGCTCATCTGCGCATAGATTATGACATAAAGATCTGCTGCTACCATAGAA 2376
Db 2169 TACTGCTTCTGCGCCACAGTGGCATGCTGCTGCTGCTGCGGCTCTGGAACCCCGACCA 2228
QY 2377 GAG 2379
Db 2229 GAG 2231

RESULT 40
HS295C061/c
LOCUS HS295C061 798 bp mRNA linear PRI 28-JAN-1999
DEFINITION H.sapiens gene from PAC 295C6, similar to rat P044.
ACCESSION AL035304
VERSION AL035304.1 GI:4200231
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Rhodes, S.
Direct Submission
Submitted (06-JAN-1999) E-mail contact: humquery@sanger.ac.uk
This sequence was generated from cDNA clones isolated using
sequence from the bacterial clone 295C6 (297876) and EST data. The
EST sequences listed match this sequence with an identity of at
least 95% between the coordinates shown.
Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1/ Sanger Centre name: dJ295C6.cl.1.
Location/Qualifiers
1. .798
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="1q24"
join(1. .169,172. .191,189. .408)
note="match: EST AA397692"

FEATURES
source
misc_feature

```



GenCore version 5.1.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: March 25, 2003, 20:15:03 ; Search time 593 Seconds  
(without alignments)  
9797.902 Million cell updates/sec

Title: US-09-781-693a-1  
Perfect score: 2580  
Sequence: 1 atgtctcgggtggtcctca.....atgaataatgagatgagaa 2580

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues  
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 120 summaries

Database : N\_Geneseq\_101002.\*

```

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2573.6	99.8	3189	22	AAF58654
2	2565.8	99.4	3208	21	AAC59068
3	2508.4	97.2	3022	22	AAF59642
4	2332.6	90.4	3420	22	AAF58655
5	2140	82.9	2448	22	AAF26955
6	1542.4	59.8	1942	24	ABK35809
7	1446.4	56.1	1952	24	ABK33744
8	1067.4	41.4	1726	22	AAH31776
9	798.4	30.9	1012	22	AAS31170

Human Repr-EN-1.0  
Human secreted pro  
Human cell cycle a  
Human IBI cDNA. H  
Human cancer assoc  
CDNA sequence #200  
Human signal trans  
Human cDNA sequenc  
Human diagnostic a

us-09-781-693a-1.rng

10	671.4	26.0	728	22	AAI97811	Human neuroblastom
11	605.6	23.5	635	21	AAA44677	Human secreted exp
12	485.4	18.8	823	22	AAH03585	Human cDNA clone (
13	328	12.7	332	22	ABA48974	Human breast cell
14	328	12.7	332	22	ABA33566	Probe #12432 for g
15	328	12.7	332	22	AAK15322	Human brain expres
16	328	12.7	332	22	AAK10553	Human bone marrow
17	328	12.7	332	22	AAI21817	Probe #11750 for g
18	328	12.7	332	22	AAI07503	Probe #7494 used t
19	287.8	11.2	710	22	AAI97812	Human genome-deriv
20	261	10.1	261	22	ABA46979	Human neuroblastom
21	261	10.1	261	22	ABA46860	Human breast cell
22	261	10.1	261	22	ABA31975	Human foetal liver
23	261	10.1	261	22	AAK13288	Probe #10441 for g
24	261	10.1	261	22	AAK39023	Human brain expres
25	261	10.1	261	22	AAI19833	Human bone marrow
26	261	10.1	261	22	AAI45028	Probe #3766 for ge
27	261	10.1	261	22	AAI05549	Probe #13714 used
28	261	10.1	261	22	ABA43873	Human genome-deriv
29	255.8	9.9	481	22	ABA43873	Human breast cell
30	255.8	9.9	481	22	ABA43873	Human genome-deriv
31	255.8	9.9	481	22	ABA43873	Human breast cell
32	255.8	9.9	481	22	ABA43873	Human genome-deriv
33	255.8	9.9	481	22	ABA43873	Human breast cell
34	255.8	9.9	481	22	ABA43873	Human genome-deriv
35	255.8	9.9	481	22	ABA43873	Human breast cell
36	255.8	9.9	481	22	ABA43873	Human genome-deriv
37	255.8	9.9	481	22	ABA43873	Human breast cell
38	248.4	9.6	423	23	ABV60730	Human genome-deriv
39	186	7.2	186	22	ABA46980	Human prostate exp
40	186	7.2	186	22	ABA46980	Human breast cell
41	186	7.2	186	22	ABA46980	Human genome-deriv
42	186	7.2	186	22	ABA46980	Human breast cell
43	186	7.2	186	22	ABA46980	Human genome-deriv
44	186	7.2	186	22	ABA46980	Human breast cell
45	186	7.2	186	22	ABA46980	Human genome-deriv
46	186	7.2	186	22	ABA46980	Human breast cell
47	175.6	6.8	178	21	AAI33108	Human genome-deriv
48	160	6.2	467	22	ABK16347	Human testicular a
49	160	6.2	467	22	ABK16347	Human testicular a
50	160	6.2	467	22	ABK16347	Human testicular a
51	160	6.2	467	22	ABK16347	Human testicular a
52	160	6.2	467	22	ABK16347	Human testicular a
53	160	6.2	467	22	ABK16347	Human testicular a
54	160	6.2	467	22	ABK16347	Human testicular a
55	160	6.2	467	22	ABK16347	Human testicular a
56	158.4	6.1	4800	22	ABK16347	Human testicular a
57	158.4	6.1	4800	22	ABK16347	Human testicular a
58	158.4	6.1	4800	22	ABK16347	Human testicular a
59	158.4	6.1	4800	22	ABK16347	Human testicular a
60	154.6	6.0	556	22	ABK16347	Human testicular a
61	141	5.5	468	22	ABK16347	Human testicular a
62	141	5.5	468	22	ABK16347	Human testicular a
63	141	5.5	468	22	ABK16347	Human testicular a
64	141	5.5	468	22	ABK16347	Human testicular a
65	141	5.5	468	22	ABK16347	Human testicular a
66	141	5.5	468	22	ABK16347	Human testicular a
67	141	5.5	468	22	ABK16347	Human testicular a
68	141	5.5	468	22	ABK16347	Human testicular a
69	70.6	2.7	2966	22	ABK16347	Human testicular a
70	70.6	2.7	2966	22	ABK16347	Human testicular a
71	70.6	2.7	2966	22	ABK16347	Human testicular a
72	70.6	2.7	2966	22	ABK16347	Human testicular a
73	70.6	2.7	2966	22	ABK16347	Human testicular a
74	68.2	2.6	630	22	ABK16347	Human testicular a
75	63.8	2.5	1887	24	ABK16347	Human testicular a
76	63.8	2.5	1887	24	ABK16347	Human testicular a
77	63.8	2.5	1887	24	ABK16347	Human testicular a
78	63.8	2.5	1887	24	ABK16347	Human testicular a
79	63.8	2.5	1887	24	ABK16347	Human testicular a
80	61	2.4	1731	24	ABK16347	Human testicular a
81	60.8	2.4	1552	21	AAK50194	Human testicular a
82	60.8	2.4	1552	21	AAK50194	Human testicular a



QY 601 CCACCAATACCATATTAACCTTCTGCTGTTGTTCTGACAGCTCAGTACCAATATATGAT 660  
 Db |||||  
 QY 776 CCACCAATACCATATTAACCTTCTGCTGTTGTTCTGACAGCTCAGTACCAATATATGAT 835  
 Db |||||  
 QY 661 CGCGAATGCTGGGCACAGAGCTACAGGGAATATACAGGCTCGAGGCTACTGGAATG 720  
 Db |||||  
 QY 836 CGCGAATGCTGGGCACAGAGCTACAGGGAATATACAGGCTCGAGGCTACTGGAATG 895  
 Db |||||  
 QY 721 GTTCCCGTTTATTCCTCCCATCTTAATAAAGCTCTGAGAGTGACATCTCTGTGT 780  
 Db |||||  
 QY 896 GTTCCCGTTTATTCCTCCCATCTTAATAAAGCTCTGAGAGTGACATCTCTGTGT 955  
 Db |||||  
 QY 781 TACAGTGAAGATGGTCAAGAGATTCCTGTTAGTTACTCTTCAGATATACATATATCTTTT 840  
 Db |||||  
 QY 956 TACAGTGAAGATGGTCAAGAGATTCCTGTTAGTTACTCTTCAGATATACATATATCTTTT 1015  
 Db |||||  
 QY 841 GACCCGAAGATGATACAGAGCTTAAACTCTCTCTGCGGAGAGAGAGAGAA 900  
 Db |||||  
 QY 1016 GACCCGAAGATGATACAGAGCTTAAACTCTCTCTGCGGAGAGAGAGAGAA 1075  
 Db |||||  
 QY 901 GAGTTGCGAACAACCACTTAAGCTTTAGAGCTTCGTGTTGATTTGCTGATCTGCTGGA 960  
 Db |||||  
 QY 1076 GAGTTGCGAACAACCACTTAAGCTTTAGAGCTTCGTGTTGATTTGCTGATCTGCTGGA 1135  
 Db |||||  
 QY 961 CCCAGAGCAAGCGGAGAGTGAACGAGAGATGGAGCAGAGTCCCAATGCTGGA 1200  
 Db |||||  
 QY 1136 CCCAGAGCAAGCGGAGAGTGAACGAGAGATGGAGCAGAGTCCCAATGCTGGA 1260  
 Db |||||  
 QY 1021 TTGATGAGAGATGCTGATATGTTATCAAGATGTTTGAAGAGCAAGTCAAGTGGAA 1315  
 Db |||||  
 QY 1196 TTGATGAGAGATGCTGATATGTTATCAAGATGTTTGAAGAGCAAGTCAAGTGGAA 1400  
 Db |||||  
 QY 1081 CAAAGCAATAGAGAGAGAGATTCGACCCAGAGTGGAGCAAGTCAAGTGGAA 1455  
 Db |||||  
 QY 1256 CAAAGCAATAGAGAGAGATTCGACCCAGAGTGGAGCAAGTCAAGTGGAA 1510  
 Db |||||  
 QY 1141 TCAACTCTTCTTACGCTCCATCAAGTCTGATTTGAGAGTGAAGTGAAGTGAAGTGAAG 1565  
 Db |||||  
 QY 1316 TCAACTCTTCTTACGCTCCATCAAGTCTGATTTGAGAGTGAAGTGAAGTGAAGTGAAG 1620  
 Db |||||  
 QY 1201 TPAGATCTCCAGCTGAACAAATTTCTCAGCTTCTACATCTCTACATCTCTACATCTCTAG 1675  
 Db |||||  
 QY 1376 TPAGATCTCCAGCTGAACAAATTTCTCAGCTTCTACATCTCTACATCTCTACATCTCTAG 1730  
 Db |||||  
 QY 1261 GCTATTGCAATCATCTCCACAGAAAGCCCTCATCTACTCTCTTGTATCTCTCTCA 1785  
 Db |||||  
 QY 1436 GCTATTGCAATCATCTCCACAGAAAGCCCTCATCTACTCTCTTGTATCTCTCTCA 1840  
 Db |||||  
 QY 1321 GACAGTGAACAAAGGAGTCTGTTGAGGATCTGGACACACACATCATCATCTCTCTCTCA 1895  
 Db |||||  
 QY 1496 GATAGTGAACAAAGGAGTCTGTTGAGGATCTGGACACACACATCATCATCTCTCTCTCA 1950  
 Db |||||  
 QY 1381 AACAAATGAAGAGTGAAGCCCAACAGGAGAGTGAACAGTGAACAGTGAACAGTGAACAG 2005  
 Db |||||  
 QY 1556 AACAAATGAAGAGTGAAGCCCAACAGGAGAGTGAACAGTGAACAGTGAACAGTGAACAG 2060  
 Db |||||  
 QY 1441 TACAGACAGAGAGAGTGAACAGTGAACAGTGAACAGTGAACAGTGAACAGTGAACAGTGAACAG 2115  
 Db |||||  
 QY 1616 TACAGACAGAGAGAGTGAACAGTGAACAGTGAACAGTGAACAGTGAACAGTGAACAGTGAACAG 2170  
 Db |||||  
 QY 1501 AGTATAGCATCAAGTCTGAGAGGATTTGGAGGATTTGAGGATTTGAGGATTTGAGGATTTGAGGAG 2225  
 Db |||||  
 QY 1676 AGTATAGCATCAAGTCTGAGAGGATTTGGAGGATTTGAGGATTTGAGGATTTGAGGATTTGAGGAG 2280  
 Db |||||  
 QY 1561 TCTTTCTGCTCCACAGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 2335  
 Db |||||  
 QY 1736 TCTTTCTGCTCCACAGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 2390  
 Db |||||  
 QY 1621 GAAGATCATCAGAGATGAGCAAAATATCAGAGAGAGTATCTGCAAGAAACCCAGTT 2445  
 Db |||||  
 QY 1796 GAAGATCATCAGAGATGAGCAAAATATCAGAGAGAGTATCTGCAAGAAACCCAGTT 2500  
 Db |||||

QY 1681 GAGAACCAATATCAATATATCAACCAATCAAGTATCAAGCCAAAGCCATTTGATTTCCAAC 1740  
 Db |||||  
 QY 1856 GAGAACCAATATCAATATATCAACCAATCAAGTATCAAGCCAAAGCCATTTGATTTCCAAC 1915  
 Db |||||  
 QY 1741 TCAGGAGAAAGAAATGACCTCAATCTTGTGCTGCTTGTGGGTTCCAGAGAAATCTGCT 1800  
 Db |||||  
 QY 1916 TCAGGAGAAAGAAATGACCTCAATCTTGTGCTGCTTGTGGGTTCCAGAGAAATCTGCT 1975  
 Db |||||  
 QY 1801 TCATCTCAAAAGGCGAAGAACCAAGAACTTCAGATCAAGTACAGTACAGTACAGTACAGTACAG 2035  
 Db |||||  
 QY 1976 TCATCTCAAAAGGCGAAGAACCAAGAACTTCAGATCAAGTACAGTACAGTACAGTACAGTACAG 2100  
 Db |||||  
 QY 1861 AATGAAATTAACCACTCACTGAGCTCTAGTTCCAAAGCAAGAACCACTGAGCTTCACT 2160  
 Db |||||  
 QY 2036 AATGAAATTAACCACTCACTGAGCTCTAGTTCCAAAGCAAGAACCACTGAGCTTCACT 2225  
 Db |||||  
 QY 1921 CATGAAGAAACATCCACAGGAGCTCTCTCTTCAAGGACACAGATGACAGTATGATGAC 2280  
 Db |||||  
 QY 2096 CATGAAGAAACATCCACAGGAGCTCTCTCTTCAAGGACACAGATGACAGTATGATGAC 2340  
 Db |||||  
 QY 1981 CCAGTCTGATCCCAAGGCTTCTTCAAGGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2400  
 Db |||||  
 QY 2041 GCGGCTATTCAGAGGCTTCTTCAAGGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2460  
 Db |||||  
 QY 2216 GCGGCTATTCAGAGGCTTCTTCAAGGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2520  
 Db |||||  
 QY 2101 ACTTTGAACATTAAGAGGCGCTAGTAAAGTGTATTAAGGCGCATCGCAACTTCCAGG 2580  
 Db |||||  
 QY 2276 ACTTTGAACATTAAGAGGCGCTAGTAAAGTGTATTAAGGCGCATCGCAACTTCCAGG 2640  
 Db |||||  
 QY 2161 ACAATGATTAAGAGGCGCTAGTAAAGTGTATTAAGGCGCATCGCAACTTCCAGG 2700  
 Db |||||  
 QY 2336 ACAATGATTAAGAGGCGCTAGTAAAGTGTATTAAGGCGCATCGCAACTTCCAGG 2760  
 Db |||||  
 QY 2221 GCGCACATTTCTATCTGGGATCGGCACACTCTGAGCAATTTGATGCTTCTGGAAGCTGAT 2820  
 Db |||||  
 QY 2396 GCGCACATTTCTATCTGGGATCGGCACACTCTGAGCAATTTGATGCTTCTGGAAGCTGAT 2880  
 Db |||||  
 QY 2281 AATCATGTTGTTAACTGCTGCGCACACTCTGAGCAATTTGATGCTTCTGGAAGCTGAT 2940  
 Db |||||  
 QY 2456 AATCATGTTGTTAACTGCTGCGCACACTCTGAGCAATTTGATGCTTCTGGAAGCTGAT 3000  
 Db |||||  
 QY 2341 ATAGATTATGACATTAAGAGTCTGCTCACCATTAGAAGAGTCAAGGATTTTAAACCGAAA 3060  
 Db |||||  
 QY 2516 ATAGATTATGACATTAAGAGTCTGCTCACCATTAGAAGAGTCAAGGATTTTAAACCGAAA 3120  
 Db |||||  
 QY 2401 CTTGCTGATGAAGTATACTCGAAACGAACTCTGGAAGAACTAGAAACCACTTAT 3180  
 Db |||||  
 QY 2576 CTTGCTGATGAAGTATACTCGAAACGAACTCTGGAAGAACTAGAAACCACTTAT 3240  
 Db |||||  
 QY 2461 ACAGTTCCAGCTCTTCTCATGTTGAGGATTTGCTTCACTTAACTATCCAGAGTAC 3300  
 Db |||||  
 QY 2636 ACAGTTCCAGCTCTTCTCATGTTGAGGATTTGCTTCACTTAACTATCCAGAGTAC 3360  
 Db |||||  
 QY 2521 CGSTTGGAGGTTGACAGATCAAGAGGCTCTGCTCAAGAGATGAAATGAGGATGAGGAA 3420  
 Db |||||  
 QY 2696 CGSTTGGAGGTTGACAGATCAAGAGGCTCTGCTCAAGAGATGAAATGAGGATGAGGAA 3480  
 Db |||||

## RESULT 2

AAC59068

ID AAC59068 standard; cdna; 3208 BP.

XX AAC59068;

AC AAC59068;

DT 02-FEB-2001 (first entry)

XX Human secreted protein coding sequence SEQ ID NO: 30.

XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;

XX antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;

KW

vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;  
 cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
 secreted protein; ss.

neurological disease; infection; human; secreted protein

Homo sapiens.

WO200055171-A1.

21-SEP-2000.

09-MAR-2000; 2000WO-US06043.

12-MAR-1999; 99US-0124146.

23-NOV-1999; 99US-0167061.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Ruben SM, Komatsoulis G;

WPI: 2000-638174/61.

P-PSDB; AAB27926.

Isolated nucleic acid molecule encoding a human secreted protein is used in preventing, treating or ameliorating a medical condition -

Claim 1: Page 344-345; 438pp; English.

The invention relates to the isolation of genes AA59049-A59098 encoding 50 human secreted proteins AB27907-827956. The genes can be used to generate fusion proteins by linking to the gene for the human immunoglobulin G Fc portion (SEQID1) for increasing the stability of the fusion protein as compared to the human protein only. The genes and proteins are useful for preventing, ameliorating or treating medical conditions, e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer, and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemia and wound healing; (d) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections.

CC and parasitic infections.

XX

SQ Sequence 3208 BP; 975 A; 667 C; 739 G; 827 T; 0 other;

Query Match 99.4%; Score 2565.8; DB 21; Length 3208;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 2578; Conservative 0; Mismatches 2; Indels 1; Gaps

[illegible]

44 / 1103

QY	301	AAGTCTTACCTGTGACAATGATAAAGATGGTAATTTTCTTCCTGCTCGAGATGGAGTAATA	566
Db	507	AAGTCTTACCTGTGACAATGATAAAGATGGTAATTTTCTTCCTGCTCGAGATGGAGTAATA	
QY	361	TTTTATACCACCGTTGAGCAAGATGCAGAAACCCACAGACAATG-CCAATTACGTGTCA	419
Db	567	TTTTATACCACCGTTGAGCAAGATGCAGAAACCCACAGACAATGCCAATTACGTGTCA	626
QY	420	TTATGGAAGTACTTTATGAGATATGACTGTACCAATGACCCCTTACACATTTCTCTCTTG	479
Db	627	TTATGGAAGTACTTTATGAGATATGACTGTACCAATGACCCCTTACACATTTCTCTCTTG	586
QY	480	TGGTGAAGTGGAACTGTTAGGTGGTTTGTATACAGCATCAAACCTAGCTGCACAAAAGA	539
Db	687	TGGTGAAGTGGAACTGTTAGGTGGTTTGTATACAGCATCAAACCTAGCTGCACAAAAGA	746
QY	540	AGATTGTAAGATGATATTTAAATAAGTTCGACGTGCTGCCAGCTCTGTTGCTATTTG	599
Db	747	AGATTGTAAGATGATATTTAAATAAGTTCGACGTGCTGCCAGCTCTGTTGCTATTTG	806
QY	600	CCCACCAATACCAATATTAACCTTGTGTTGTTCTGACAGCTCAGTACGAATATATGA	659
Db	807	CCCACCAATACCAATATTAACCTTGTGTTGTTCTGACAGCTCAGTACGAATATATGA	866
QY	660	TCGGCGAATGCTGGCACAAAGCTACAGGGAATATGACGTCGAGGACCTACTGGAAT	719
Db	867	TCGGCGAATGCTGGCACAAAGCTACAGGGAATATGACGTCGAGGACCTACTGGAAT	926
QY	720	GTTTGCCTGTTTATCTCTCCCATCTTAATAAGTCTCTGAGAGTCGACATCTGTG	779
Db	927	GTTTGCCTGTTTATCTCTCCCATCTTAATAAGTCTCTGAGAGTCGACATCTGTG	839
QY	780	TTACAGTCAAGATGGTCAAGAGATTCGTGTAGTTACTCTTCAGATTACATATATCTTTT	1046
Db	987	TTACAGTCAAGATGGTCAAGAGATTCGTGTAGTTACTCTTCAGATTACATATATCTTTT	899
QY	840	TGACCGGAAGATGATACAGACAGAGAACTTAAAACTCTTTCGGGAAGAGAAGAGA	1106
Db	1047	TGACCGGAAGATGATACAGACAGAGAACTTAAAACTCTTTCGGGAAGAGAAGAGA	1166
QY	900	AGAGTTGCGCAACACACAGTTAAGCGTTTGAGACTTCGTGGTGATGGTCAGACTGCG	959
Db	1107	AGAGTTGCGCAACACACAGTTAAGCGTTTGAGACTTCGTGGTGATGGTCAGACTGCG	1166
QY	960	ACCCAGAGCAAGCCGAGAGTGAACGAGNACAGATGGAGAGCAGAGTCCCAATGTGTC	1019
Db	1167	ACCCAGAGCAAGCCGAGAGTGAACGAGNACAGATGGAGAGCAGAGTCCCAATGTGTC	1226
QY	1020	ATTGATGCAGAGATGCTGATATGTTTACAAGATGGTTTGAGAGACAGTGAAGTTCG	1079
Db	1227	ATTGATGCAGAGATGCTGATATGTTTACAAGATGGTTTGAGAGACAGTGAAGTTCG	1286
QY	1080	ACAAAGCAATAGAGCAGCAAGATCTCGACCCAGAGTGGAAACAGTCAATCAGATAT	1139
Db	1287	ACAAAGCAATAGAGCAGCAAGATCTCGACCCAGAGTGGAAACAGTCAATCAGATAT	1346
QY	1140	TTCAACTCTTCTACGGTCCCATCAAGTCTCTGATTTGGAAGTGAAGTGAAGTGAAGTGA	1199
Db	1347	TTCAACTCTTCTACGGTCCCATCAAGTCTCTGATTTGGAAGTGAAGTGAAGTGAAGTGA	1406
QY	1200	AGTAGTACTCCAGCTGAACAATTTCTTCAGCCTTCTACATCTTCAATGTCAGTCTCA	1259
Db	1407	AGTAGTACTCCAGCTGAACAATTTCTTCAGCCTTCTACATCTTCAATGTCAGTCTCA	1466
QY	1260	GGCTCATTCGACATCATCTCCACAGAAAGCCCTCATTTCTTACCTTTCGATCTCTCC	1319
Db	1467	GGCTCATTCGACATCATCTCCACAGAAAGCCCTCATTTCTTACCTTTCGATCTCTCC	1526
QY	1320	AGACAGTCAACAAAGGCGAGTGTGTGAGGATCTGGACACACACATCATCAGTCTGA	1379
Db	1527	AGATAGTCAACAAAGGCGAGTGTGTGAGGATCTGGACACACACATCATCAGTCTGA	1586
QY	1380	TAACATTAATGAAGTGTAGCCCCAACACACAGGAGTGAACACAGTCTTAAAGTTTGA	1439

Db	1587	TAACAATAATGAAAGCTGAGCCCCAACCAGGACAGGTGAACCAAGTTTAAAGTTTGA	1646
QY	1440	CTACAGCAGAGGAACAACACTACAAGCACAAATAAACTGAACTTTACAGATGAATGGAG	1499
Db	1647	CTACAGCAGAGGAACAACACTACAAGCACAAATAAACTGAACTTTACAGATGAATGGAG	1706
QY	1500	CAGTATAGCATCAAGTCTTAGAGGAATTTGGAGCCATTGCAATCTGAGGGTCAGGAGGA	1559
Db	1707	CAGTATAGCATCAAGTCTTAGAGGAATTTGGAGCCATTGCAATCTGAGGGTCAGGAGGA	1766
QY	1560	ATCTTTGCTCCACAGAGCTCAGTGCACACCAGGAAGGACAGTGAACAACAAAGCTCC	1619
Db	1767	ATCTTTGCTCCACAGAGCTCAGTGCACACCAGGAAGGACAGTGAACAACAAAGCTCC	1826
QY	1620	TGAAGAAATCATCAGAGGATGTGACAAATATACGGAAGGAGTATCTGCAGAAACCCAGT	1679
Db	1827	TGAAGAAATCATCAGAGGATGTGACAAATATACGGAAGGAGTATCTGCAGAAACCCAGT	1886
QY	1680	TGAGAACCATATCAATATAACACAATCAGATAAGTTTCAGGCCAAGCCATTGGATTCCAA	1739
Db	1887	TGAGAACCATATCAATATAACACAATCAGATAAGTTTCAGGCCAAGCCATTGGATTCCAA	1946
QY	1740	CTCAGGAGAAGAAATGACCTCAATCTTGATCGCTCTTGTGGGGTTCAGAGAATCTGC	1799
Db	1947	CTCAGGAGAAGAAATGACCTCAATCTTGATCGCTCTTGTGGGGTTCAGAGAATCTGC	2006
QY	1800	TTCATCTGAAAAAGCCAGGAACACAGAAATCTCAGATCAGACTAGACTGAGAGTCTAC	1859
Db	2007	TTCATCTGAAAAAGCCAGGAACACAGAAATCTCAGATCAGACTAGACTGAGAGTCTAC	2066
QY	1860	CAATGAAATTAACACCAATCTTGAGCCTCAGTTCCTCCAAACAGAACCACTGGGCTTCAGC	1919
Db	2067	CAATGAAATTAACACCAATCTTGAGCCTCAGTTCCTCCAAACAGAACCACTGGGCTTCAGC	2126
QY	1920	TCATGAAGAAACATCCACAGGAGCTCTGCTCTTCAGGACACACATGACATGATGATGA	1979
Db	2127	TCATGAAGAAACATCCACAGGAGCTCTGCTCTTCAGGACACACATGACATGATGATGA	2186
QY	1980	CCAGTCTGATCCAGTGCAGATATCGAGCAGGACCTGGTGATAGAGCTCTGCTGT	2039
Db	2187	CCAGTCTGATCCAGTGCAGATATCGAGCAGGACCTGGTGATAGAGCTCTGCTGT	2246
QY	2040	TGCCCGTATTTCAGGAGTTCTTCAGACGAGAAAGAAAGAAATGGAAGAAATTTGA	2099
Db	2247	TGCCCGTATTTCAGGAGTTCTTCAGACGAGAAAGAAAGAAATGGAAGAAATTTGA	2306
QY	2100	TACTTTGAACATTTAGAGCGCGCTAGTAAAAATGGTTTATAAGGCCATCGCACTCCAG	2159
Db	2307	TACTTTGAACATTTAGAGCGCGCTAGTAAAAATGGTTTATAAGGCCATCGCACTCCAG	2366
QY	2160	GACATGATAAAGAGCCAAATCTCGGGTGCTACTTTGTAATGAGTGGTCTGACTG	2219
Db	2367	GACATGATAAAGAGCCAAATCTCGGGTGCTACTTTGTAATGAGTGGTCTGACTG	2426
QY	2220	TGGCCCATTTTCATCTGGATCGGCACACTGCTGAGCATTTGATGCTCTCTGGAAGCTGA	2279
Db	2427	TGGCCCATTTTCATCTGGATCGGCACACTGCTGAGCATTTGATGCTCTCTGGAAGCTGA	2486
QY	2280	TAATCATGTGGTAAACTGGCTGCAGCCACATCGTTTGACCCAAATTTAGCCCTCATCTGG	2339
Db	2487	TAATCATGTGGTAAACTGGCTGCAGCCACATCGTTTGACCCAAATTTAGCCCTCATCTGG	2546
QY	2340	CATAGATTTAGCATAAAGATCTGGTTCACCATTAGAAGTCAAGGATTTTAAACCGAAA	2399
Db	2547	CATAGATTTAGCATAAAGATCTGGTTCACCATTAGAAGTCAAGGATTTTAAACCGAAA	2606
QY	2400	ACTTGTCTGATGAAGTTTAACTCGAAGCAACTCATGCTGGAAGAAATAGAAAACACAT	2459
Db	2607	ACTTGTCTGATGAAGTTTAACTCGAAGCAACTCATGCTGGAAGAAATAGAAAACACAT	2666
QY	2460	TACAGTTCAGGCTCTTTTCATCTTCAGGATGTGCGCTTCATTAATCATATTCGAGCTGA	2519

Db 2567 TACAGTTCCAGCCCTTTTCATGTTGAGAGATGTTGGCTTCACTTAATCATATCCGAGCTCA 2726

Qy 2520 CCGGTTGGAGGGTGACAGATCAGAAGGCTCTGGTCAAGAGAATGAAAATGAGGATGAGGA 2579  
|||||

Db 2727 CCGGTTGGAGGGTGACAGATCAGAAGGCTCTGGTCAAGAGAATGAAAATGAGGATGAGGA 2786  
|||||

Qy 2580 A 2580  
|

Db 2787 A 2787

RESULT 3

AAF59642

ID AAF59642 standard; cDNA; 3022 BP.

XX

AC AAF59642;

XX

DT XX

XX 24-APR-2001 (first entry)

XX Human cell cycle and proliferation protein CCYPR-53 cDNA, SEQ ID NO:107.

XX

KW Cell cycle and proliferation protein; CCYPR; human; agonist;

KW antagonist; gene therapy; detection; gene therapy;

KW transgenic animal disease model; immune disorder;

KW developmental disorder; cell signalling disorder;

KW cell proliferative disorder; cancer; tumour; anaemia; epilepsy;

KW arteriosclerosis; asthma; allergy; diabetes mellitus;

KW menstrual cycle disorder; bacterial infection; ss.

XX

OS Homo sapiens.

XX

XX WO200107471-A2.

PN

XX 01-FEB-2001.

XX

XX 21-JUL-2000; 2000WO-US19948.

XX

XX 21-JUL-1999; 99US-0145075.

PR

PR 08-SEP-1999; 99US-0153129.

PR

XX 10-NOV-1999; 99US-0164647.

XX

PA (INCY-) INCYTE GENOMICS INC.

XX

PI Hillman JL, Lal P, Tang YT, Yue H, Au-Young J, Bandman O;

PI Azimzai Y, Yang J, Lu DAM, Baughn MR, Patterson C, Shah P;

XX

DR WPI; 2001-112727/12.

DR

XX P-PSDB; AAB60505.

XX

Human cell cycle and proliferation proteins and polynucleotides are used to treat, diagnose and prevent immune, developmental and cell signalling disorders and cell proliferative disorders including cancer - Claim 5; Page 204; 205pp; English.

PS

XX Sequences AAF59590-AAF59643 represent cDNAs encoding 54 human cell cycle and proliferation proteins (CCyPR), AAB60453-AAB60506. CCyPR and agonists of CCyPR are used to treat diseases or conditions associated with decreased expression of functional CCyPR, while CCyPR antagonists are used to treat diseases or conditions associated with overexpression of functional CCyPR. Monoclonal or polyclonal antibodies to CCyPR may be used in enzyme-linked immunosorbent assays (ELISA) or radioimmunoassays to detect CCyPR. CCyPR itself may be used to detect compounds e.g., antibodies, oligonucleotides and proteins (receptors) that specifically bind to CCyPR, and in drug screening methods to identify compounds that modulate the activity of CCyPR. CCyPR nucleotides can be used to generate transgenic animal models of human disease, and can be used in gene therapy in target cells with genetic abnormalities with respect to the expression of CCyPR for the treatment or prevention of a disorder associated with CCyPR. Diseases which can be diagnosed, treated and prevented using CCyPR proteins, nucleic acids, agonists or antagonists include immune, developmental and cell signalling disorders, and cell proliferative

disorders including cancer. Specific examples of these disorders include anaemia, epilepsy, arteriosclerosis, asthma, cancer, allergies, diabetes mellitus, disorders of the menstrual cycle and infections caused by bacteria.

XX 027 A: 611 C: 674 G: 800 T: 0 other;

```
Query Match
97.2%; Score 2508.4; DB 22; Length 3022;
```

Query Match: 97.78; Pred. No. 0;  
Best Local Similarity

Matches 2579; Conservative 0; Mismatches 1; Indels 00; Gaps 1

QY  
1 ATGTC TCGGGTGGCTCCTACCCACACCTGTTGTGGGACGTTGAGGAAAGGTTCCCTCGG 60

[illegible]

**D**b      119 ATGTCCTCGGGGTGGCTTGCACACCAGATGATGCCAAGCAAGAAAGA  
C

61 CTGGAGGACCCGTCCTCCGGCTGCGGAGTCGCTACCTGGGAAGAAGAGAAATTATCCAAGA 120

[illegible]

db 179 CTGGAGGACCGTCCCGCTGCGGAGTCGCTACCTGGGAAGAGAAATTATCCAAGA 230

... - - - - - mcmccmcmcmccmcttatacaatctgttggaat 180

QY  
121 TTTAAACTTGAAGCAACCCCTTAATGTGCAATGGATTGCTTATTCAGC

220 TTTTCTTCTGCAACCTTAA TGTGCATGATGGTTGTGTTAATACAATCTGTTGGAAT 298

[illegible]

181 GACACTGGAGAAATATTTTATCTGGCTCAGATGACACCAATTAGTAAATTAGTAACT 240

358

**Db**

... TGGACCAACAATTCTGAGGGCACCGAGCAACATATTAGTGCA 300

QY  
241 TACAGCAGATAGGTTTGGCTACGATTTCCTTCACTT

nb 359 TACAGCAGAAAGGTTTTCACAACAATTCTTCAGGGCACCAGCAACATATTTAGTGCA 418

555 360

301 AAGTCTTACCTTGTACAAATGATAACACAGATTGTAATCCCTCTGGAGATGGAGCTTAA  
QY

**TCTTATCCCTGGAGATGGAGTAATA** 478

**Db**

03' 361 TTTTATACCAACGTTGAGCAAGATGCAGAAACCAACAGACAATGCCAATTACGTGTTCAT 420

[illegible]

Db 479 TTTTATACCAACGTTGAGCAAGATGCAGAAACCAACAGACAAATGCCAAATTACGGIGICAT

-----CAGGCTCCGAAAGACCCCTTACACTTTTCTCTCTGT 480

[illegible]

520 TATCCCACTACTCTATGACTGTACCCAATGACCCCTACACTTTTCTCTCTTGT 598

DB 339 INFOANTACTINCHONEPHEP

-----CCCCCCCCC EAC

481 GGTAAGATGGAAC TGTAGGTGTTGATACACGCATCAAACTAGC TGCACAAAGHA 340

|||||CTAGCTGCACAAAAGAA 658

**Db** 599 GGTGAAGATGGAAC TGT TAGGTGGT TTGATACACGCATCAATTCACCTCCATTCTTT

5' - C A G T C G A C G T G C T G C C A C G T C T G T T G C T A T T G C 600



QY 1981 CCAGTCCTGATCCAGGTCAGGATATCGAGCAGGACCTGGTATAGACGCTCTCTGTT 2040  
 Db |||||||  
 QY 2159 CCAGTCCTGATCCAGGTCAGGATATCGAGCAGGACCTGGTATAGACGCTCTCTGTT 2218  
 Db |||||||  
 QY 2041 GCCGTAATCAGGAGTCTTCAGACGGAGAGAAAGAAAGAAAGAAATGGAATTTGGAT 2100  
 Db |||||||  
 QY 2101 ACTTGAACATTAGAAGCCCTAGTAAATGTTTATAGAGCCATCGCAATCCAGG 2160  
 Db |||||||  
 QY 2279 ACTTGAACATTAGAAGCCCTAGTAAATGTTTATAGAGCCATCGCAATCCAGG 2338  
 Db |||||||  
 QY 2161 ACAATGATAAAGAGCAATTTCTGGGGTCTAACTTTTGAATGAGTGTCTGACTGT 2220  
 Db |||||||  
 QY 2339 ACAATGATAAAGAGCAATTTCTGGGGTCTAACTTTTGAATGAGTGTCTGACTGT 2398  
 Db |||||||  
 QY 2221 GGCACATTTTCACTCTGGGATCGGCACACTGCTGAGCATTTGATGCTTCTGGAAGCTGAT 2280  
 Db |||||||  
 QY 2399 GGCACATTTTCACTCTGGGATCGGCACACTGCTGAGCATTTGATGCTTCTGGAAGCTGAT 2458  
 Db |||||||  
 QY 2281 AATCATGTGTAAACTGCTGAGCCACATCCGTTTACCCAAATTTAGCCTCATCTGGC 2340  
 Db |||||||  
 QY 2459 AATCATGTGTAAACTGCTGAGCCACATCCGTTTACCCAAATTTAGCCTCATCTGGC 2518  
 Db |||||||  
 QY 2341 ATAGATTATGACATAAAGATCTGCTCACCATTAGAAGATCAAGGATTTTAAACCGAAA 2400  
 Db |||||||  
 QY 2519 ATAGATTATGACATAAAGATCTGCTCACCATTAGAAGATCAAGGATTTTAAACCGAAA 2578  
 Db |||||||  
 QY 2401 CTTGCTGATGAAGTTATTAACCTGAAACGAACTCATGCTGGAAGAACTAGAAACACCAT 2460  
 Db |||||||  
 QY 2579 CTTGCTGATGAAGTTATTAACCTGAAACGAACTCATGCTGGAAGAACTAGAAACACCAT 2638  
 Db |||||||  
 QY 2461 ACAGTTCAGCCTCTTTCATGTTGAGGATGTTGGCTTCACTTAATCATATCCGAGCTGAC 2520  
 Db |||||||  
 QY 2639 ACAGTTCAGCCTCTTTCATGTTGAGGATGTTGGCTTCACTTAATCATATCCGAGCTGAC 2698  
 Db |||||||  
 QY 2521 CGTTGGAGGTGACAGATCAGAGGCTCTGCTCAAGAGAAATGAAATCAGGATGAGGAA 2580  
 Db |||||||  
 QY 2699 CGTTGGAGGTGACAGATCAGAGGCTCTGCTCAAGAGAAATGAAATCAGGATGAGGAA 2758  
 Db |||||||

## RESULT 4

AAF58655  
 ID AAF58655 standard; cdna; 3420 BP.

XX AC AAF58655;

XX DX 27-APR-2001 (first entry)

XX DE Human IB1 cdna.

XX KW Human; Repro-EN-1.0; IB1; cytotstatic; breast cancer; uterine cancer;  
 XX KW prostate cancer; epitope; ss.

XX OS Homo sapiens.

XX XX WO200107616-A1.

XX PN 01-FEB-2001.

XX XX 10-MAR-2000; 2000WO-US06742.

XX PR 22-JUL-1999; 99US-0359084.

XX PR 30-JUL-1999; 99WO-US17284.

XX PR 23-NOV-1999; 99US-0447399.

XX PA (DIAG-) DIAGNOSTIC PROD CORP.

XX PI El Shami AS, Menon SN, French CK;

XX XX WPI; 2001-182795/18.

XX DR

XX XX

PT

PT

PT

XX

PS

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

PT

PT

PT

XX

PS

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

PT

PT

PT

XX

PS

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

PT

PT

PT

XX

PS

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

|||||  
Db 776 CCACCAATACCATATACCTTGGTCTGAGAGCTCAGTACGAAATATATGAT 835  
QY 661 CGCGGAATGCTGGGCACAAAGAGCTACAGGGAATATGAGCTCGAGGAGTCTGGAATG 720  
Db 836 CGCGGAATGCTGGGCACAAAGAGCTACAGGGAATATGAGCTCGAGGAGTCTGGAATG 895  
QY 721 GTTGCCCGTTTATCTTCCCATCTTAAATATAAGTCTGAGAGTCAATCTCTGTGT 780  
Db 896 GTTGCCCGTTTATCTTCCCATCTTAAATATAAGTCTGAGAGTCAATCTCTGTGT 955  
QY 781 TACAGTGAAGATGGTCAAGAGATCTCGTTAGTACTCTTACAGTACATATATCTTTT 840  
Db 956 TACAGTGAAGATGGTCAAGAGATCTCGTTAGTACTCTTACAGTACATATATCTTTT 1015  
QY 841 GACCCGAAGAGATGATACAGCAGAGAACTTAAACTCTTCTGCGGAAGAGAGAGAA 900  
Db 1016 GACCCGAAGAGATGATACAGCAGAGAACTTAAACTCTTCTGCGGAAGAGAGAGAA 1075  
QY 901 GAGTTGCGCAACACCCAGTTAAGCGTTTGAAGTCTGCTGGTGTGATGGTCACTGGA 960  
Db 1076 GAGTTGCGCAACACCCAGTTAAGCGTTTGAAGTCTGCTGGTGTGATGGTCACTGGA 1135  
QY 961 CCCAGCAAGCGCGGAGTGAACGAGACGAGATGGAGCAGAGTCCCAATGTGTCA 1020  
Db 1136 CCCAGCAAGCGCGGAGTGAACGAGACGAGATGGAGCAGAGTCCCAATGTGTCA 1195  
QY 1021 TTGATGACAGAGATGCTGTATATGTTATCAAGATGGTTGAAGCAAGTGAAGTTGCA 1080  
Db 1196 TTGATGACAGAGATGCTGTATATGTTATCAAGATGGTTGAAGCAAGTGAAGTTGCA 1255  
QY 1081 CAAGCAATAGACAGCAGGAGATCTCGACCGAGGTGGAAAGTCAATCAGATATT 1140  
Db 1256 CAAGCAATAGACAGCAGGAGATCTCGACCGAGGTGGAAAGTCAATCAGATATT 1315  
QY 1141 TCAACTCTTCTACGGTCCCATCAAGTCTGATTTGAAGTGAAGTGAAGTGAAGTGAAG 1200  
Db 1316 TCAACTCTTCTACGGTCCCATCAAGTCTGATTTGAAGTGAAGTGAAGTGAAGTGAAG 1375  
QY 1201 GTAGATCTCAGCTGAACATTTCTCAGCTCTTACATCTCTACATCTCAGCTCAG 1260  
Db 1376 GTAGATCTCAGCTGAACATTTCTCAGCTCTTACATCTCTACATCTCAGCTCAG 1435  
QY 1261 GCTCATCGACATATCTCCACAGAAAGCCCTCATCTACTCTTCTGCTATCTTCTCCA 1320  
Db 1436 GCTCATCGACATATCTCCACAGAAAGCCCTCATCTACTCTTCTGCTATCTTCTCCA 1495  
QY 1321 GACAGTGAACAAAGGAGTCTGTTGAGGATCTGGACACACACATCATCAGTCT --- 1377  
Db 1496 GATAGTGAACAAAGGAGTCTGTTGAGGATCTGGACACACACATCATCAGTCTGAA 1555  
QY 1378 ----- 1377  
Db 1556 TTTTAAAGGGGCTGAGATAGTTTGTCTGTAAGCGCTGCAACAACTGAGGCTTAAG 1615  
QY 1378 ----- 1377  
Db 1616 AAGGCTGAGCAGCAGGAGCAGAGCTAGCTGCACATACCACGAGCCCTCCACT 1675  
QY 1378 ----- 1377  
Db 1676 TCTGATCAGTCTTCTCATGAGGCTCTTACAGGACCCCTCATGCTTCAGATCTCTCTCT 1735  
QY 1378 ----- -CATACCAATAAT 1389  
Db 1736 TCTGTGTTAAACAAAGCTCGGATCCATGTCTACTTGAAGCAACAGGATACCAATAT 1795  
QY 1390 GAAAGCTGAGCCCAACCAAGGAGCAGAGTGAACCAAGTTTAAAGTTTGCATACAGCACA 1449  
Db 1796 GAAAGCTGAGCCCAACCAAGGAGCAGAGTGAACCAAGTTTAAAGTTTGCATACAGCACA 1855  
QY 1450 GAAGGAACAACTACAGCACAATAAACTGAACTTTACAGATGAATGGAGCAGTATAGCA 1509  
|||||

1856 GAAGGAACAACTACAGCACAATAAACTGAACTTTACAGATGAATGGAGCAGTATAGCA 1915  
QY 1510 TCAAGTTCTAGAGGAATTTGGAGCCATTTGCAATCTGAGGCTCAGAGGAATCTTTTCGTC 1569  
Db 1916 TCAAGTTCTAGAGGAATTTGGAGCCATTTGCAATCTGAGGCTCAGAGGAATCTTTTCGTC 1975  
QY 1570 CCACAGAGCTCAGTGCACCAACCCAGAGGAGAGCTGAAACAAAAGCTCTCTGAAGATCA 1629  
Db 1976 CCACAGAGCTCAGTGCACCAACCCAGAGGAGAGCTGAAACAAAAGCTCTCTGAAGATCA 2035  
QY 1630 TCAGAGATGTGACAAATATCAGGAAGAGTATCTGCAAGAAACCCAGTTGAGAACCAT 1689  
Db 2036 TCAGAGATGTGACAAATATCAGGAAGAGTATCTGCAAGAAACCCAGTTGAGAACCAT 2095  
QY 1690 ATCAATATAACAAATCAGATTAAGTTTCAACCCAGCCATTTGATTTCCAACTCAGGAA 1749  
Db 2096 ATCAATATAACAAATCAGATTAAGTTTCAACCCAGCCATTTGATTTCCAACTCAGGAA 2155  
QY 1750 AGAATGACCTCAATCTTGTGCTCTTGTGGGTCTTCCAGAGAAATCTGCTTCTCATCGAA 1809  
Db 2156 AGAATGACCTCAATCTTGTGCTCTTGTGGGTCTTCCAGAGAAATCTGCTTCTCATCGAA 2215  
QY 1810 AAAGCAAGGAAACCAAACTTCAGATCAGACTAGCAGTGTGAGTGTCTACCAATGAAAT 1869  
Db 2216 AAAGCAAGGAAACCAAACTTCAGATCAGACTAGCAGTGTGAGTGTCTACCAATGAAAT 2275  
QY 1870 AACACCAATCTTGCCTCAGTTCACAAACAGAGCCCTGCGCTTCCAGGAAATCTGCTCATGAA 1929  
Db 2276 AACACCAATCTTGCCTCAGTTCACAAACAGAGCCCTGCGCTTCCAGGAAATCTGCTCATGAA 2335  
QY 1930 ACATCCACGAGGACTCTGCTCTTCCAGGACACAGATGATGATGATGATGATGATGATG 1989  
Db 2336 ACATCCACGAGGACTCTGCTCTTCCAGGACACAGATGATGATGATGATGATGATGATG 2395  
QY 1990 ATCCAGGTGCAAGTATCGACAGGACCTGCTGATAGACGCTCTGCTGCTGCTGCTGCT 2049  
Db 2396 ATCCAGGTGCAAGTATCGACAGGACCTGCTGATAGACGCTCTGCTGCTGCTGCTGCT 2455  
QY 2050 CAGGAGTCTTCCAGCGGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 2109  
Db 2456 CAGGAGTCTTCCAGCGGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 2515  
QY 2110 ATTAGAGGCGGCTAGTAAATGTTTAAAGGCAATCGCAACTCCAGGCAATGATA 2169  
Db 2516 ATTAGAGGCGGCTAGTAAATGTTTAAAGGCAATCGCAACTCCAGGCAATGATA 2575  
QY 2170 AAAGAGCCCAATTTCTGGGCTGCTAACTTTGATGAGTGTCTGCTGCTGCTGCTGCT 2229  
Db 2576 AAAGAGCCCAATTTCTGGGCTGCTAACTTTGATGAGTGTCTGCTGCTGCTGCTGCT 2635  
QY 2230 TTTATCTGGGATCGGCACACTGCTGAGCATTGATGCTTCTGGAAGCTGATAATCATGTG 2289  
Db 2636 TTTATCTGGGATCGGCACACTGCTGAGCATTGATGCTTCTGGAAGCTGATAATCATGTG 2695  
QY 2290 GTAACTGCTGAGCCACACATCCGTTTGAAGCCCAATTTAGCCCTCATCTGCAATGATAT 2349  
Db 2696 GTAACTGCTGAGCCACACATCCGTTTGAAGCCCAATTTAGCCCTCATCTGCAATGATAT 2755  
QY 2350 GACATAAAGATCTGGTCAACCAATTAAGAGAGTCAAGAGATTTTAAACCAAACTGCTGAT 2409  
Db 2756 GACATAAAGATCTGGTCAACCAATTAAGAGAGTCAAGAGATTTTAAACCAAACTGCTGAT 2815  
QY 2410 GAAATTAATCTGAAAGCAACTCATGCTGGAAGAACTAGAAAACCACTTACAGTTTCCA 2469  
Db 2816 GAAATTAATCTGAAAGCAACTCATGCTGGAAGAACTAGAAAACCACTTACAGTTTCCA 2875  
QY 2470 GCCTCTTCTCATGTTGAGGATGTGGCTTCACTTAATCATATCCAGGCTGACCGGTGGAG 2529  
Db 2876 GCCTCTTCTCATGTTGAGGATGTGGCTTCACTTAATCATATCCAGGCTGACCGGTGGAG 2935  
QY 2530 GGTGACAGATCAGAGGCTCTGTTCAAGAGAAATGAAATGAGGATGAGGAA 2580  
Db 2936 GGTGACAGATCAGAGGCTCTGTTCAAGAGAAATGAAATGAGGATGAGGAA 2986

## RESULT 5

AAF26955  
ID AAF26955 standard; cDNA; 2448 BP.

XX AC AAF26955;

XX DT 10-APR-2001 (first entry)

XX DE Human cancer associated antigen precursor HOM-TES-88/94/95 cDNA SEQ ID:8.

XX KW Human; cancer associated antigen precursor; cancer associated antigen;  
seminoma; HLA; human leukocyte antigen; cytostatic; gene therapy;

XX KW vaccine; cancer; ss.

XX OS Homo sapiens.

XX PN WO200100874-A2.

XX PD 04-JAN-2001.

XX PF 23-JUN-2000; 2000WO-US17207.

XX PR 30-JUN-1999; 99US-0346498.

XX PA (LUDW-) LUDWIG INST CANCER RES.

XX PI Sahin U, Tureci O, Pfreundschuh M;

XX DR WPI; 2001-112465/12.

XX PT Diagnosing a disorder characterized by expression of a human cancer  
associated antigen precursor, comprises detecting interaction of an  
agent with a nucleic acid molecule encoding the antigen precursor.

XX PS Claim 48; Page 94; 126pp; English.

XX The present invention describes a method for diagnosing a disorder  
characterised by expression of a human cancer associated antigen (CAA)  
precursor (I) coded by a NA Group 1 nucleic acid molecule (N1)  
comprising contacting the biological sample with an agent (A) that  
specifically binds to N1, (I) or its fragment, complexed with an human  
leukocyte antigen (HLA) molecule and determining the interaction between  
the agent and N1 or (I). (I) has cytostatic activity and can be used in  
a gene therapy and vaccine production. The method can be used for treating  
a subject with a condition characterised by expression of (I) in cells  
of a subject. The present sequence represents a human cancer associated  
antigen precursor cDNA sequence which is used in the exemplification of  
the present invention.

XX SQ Sequence 2448 BP; 778 A; 513 C; 551 G; 606 T; 0 other;

Query Match 82.9%; Score 2140; DB 22; Length 2448;

Best Local Similarity 95.4%; Pred. No. 0;

Matches 2303; Conservative 0; Mismatches 45; Indels 66; Gaps 7;

QY 228 AATTAGTAACTCTTACAGCAGAAAGGTTTGACAAACAATTCGTCAGGSCACCGAGCAAA 287

DB 16 AGTAAGGAGAGGTTTCTCATCAGAGGTTTGACAAACAATTCGTCAGGSCACCGAGCAAA 75

QY 288 CATATTTAGTGCAGAAAGTTTACCTTTGACAAATGATAAACAAGATTGTATCTGCTGG 347

DB 76 CATATTTAGTGCAGAAAGTTTACCTTTGACAAATGATAAACAAGATTGTATCTGCTGG 135

QY 348 AGATGAGTAATATTTATACCAAGTTGAGCAAGATGCAGAAACCAACAGACAATGCCA 407

DB 136 AGATGAGTAATATTTATACCAAGTTGAGCAAGATGCAGAAACCAACAGACAATGCCA 195

QY 408 ATTACGTGTCATTATGGAAGTACTTATGAGATTATGACTGTACCAATGACCTTTACAC 467

DB 196 ATTTACGTGTCATTATGGAAGTACTTATGAGATTATGACTGTACCAATGACCTTTACAC 255

QY 468 TTTTCTCTCTCTGTGGTGAAGA-TGGAACGTGTAGGTGGTTTGATACAGCATCAAAACTA 526  
DB 256 TTTTCTCTCTCTGTGGTGAAGAGTGAACGTGTAGGTGGTTTGATACAGCATCAAAACTA 315  
QY 527 GCTGCACAAAAGAAGATTCTAAAGATGATATTTAATTAACGTGTCAGCGTGCACAGT 586  
DB 316 GCTGCACAAAAGAAGATTCTAAAGATGATATTTAATTAACGTGTCAGCGTGCACAGT 375  
QY 587 CTGTT--GCTATTTGCCACCAATACCATATTA-CCTTCTGTGTGGTTCTTCTGACAGT 643  
DB 376 CTGTTTGTGATTTTGGCCACCAATACCATATTAACCTTGTGTGTGTGTGTGTGACAGT 435  
QY 644 CAGTACGATATATGATCGGCGAATGCTGGGCACAGAGCTACAGGGAATATTCAGAGT 703  
DB 436 CAGTACGATATATGATCGGCGAATGCTGGGCACAGAGCTACAGGGAATATTCAGAGT 495  
QY 704 GAGGACTACTGGAATGTTGCCCTTTTATTCCTTCCCATCTTATATATAGTCTGCA 763  
DB 496 GAGGACTACTGGAATGTTG--CCGTTTATTCCTTCCCATCTTATATATAGTCTGCA 554  
QY 764 GAGTACATCTCTGTGTGTACAGTGAAGTGTCAAGAGATTCCTGTAGTTACTTCTCAG 823  
DB 555 GAGTACATCTCTGTGTGTACAGTGAAGTGTCAAGAGATTCCTGTAGTTACTTCTCAG 614  
QY 824 ATTACATATATCTTTTTCACCCGGAAGATGATACAGCAGAGAACTTAAACTCCTTCTG 883  
DB 615 ATTACATATATCTTTTTCACCCGGAAGATGATACAGCAGAGAACTTAAACTCCTTCTG 674  
QY 884 CGGAAGAGAGAAAGAGTGTGGACAACCAAGTAAAGCGTTTGAAGCTTTCAGATGTTGAAG 1063  
DB 795 AGAGTCCCAATGTGTCATTTGATGCGAGAGATGCTGATATGTTATCAAGATGTTGAAG 854  
QY 1064 AAGCAAGTGAAGTTGACAAAGCAATAGAGACGAGGAAGATCTCGACCCAGAGGTGAA 1123  
DB 855 AAGCAAGTGAAGTTGACAAAGCAATAGAGACGAGGAAGATCTCGACCCAGAGGTGAA 914  
QY 1124 CAAGTCAATCAGATATTTCAACTCTTCTACGTCCTCATCAAGTCCCTGATTTGAAATGA 1183  
DB 915 CAAGTCAATCAGATATTTCAACTCTTCTACGTCCTCATCAAGTCCCTGATTTGAAATGA 974  
QY 1184 GTGAACTGCAATGGAAGTAGATCTCCAGCTGAACAATTTCTTCAGCCTTCTACATCCT 1243  
DB 975 GTGAACTGCAATGGAAGTAGATCTCCAGCTGAACAATTTCTTCAGCCTTCTACATCCT 1034  
QY 1244 CTACAATGTCAGCTCAGGCTCATTCGACATCATCTCCACAGAAAGCCCTCATTTACT 1303  
DB 1035 CTACAATGTCAGCTCAGGCTCATTCGACATCATCTCCACAGAAAGCCCTCATTTACT 1094  
QY 1304 CTTTGTATCTTCTCCAGACAGTGAACAAGGAGTCTGTGTAGGAGTCTGGACACACA 1363  
DB 1095 CTTTGTATCTTCTCCAGACAGTGAACAAGGAGTCTGTGTAGGAGTCTGGACACACA 1154  
QY 1364 CACATCATCAGTCTGATAACAATATGAAGCTGAGC----- 1401  
DB 1155 CACATCATCAGTCTGATTCTCTTCTTCTGTGTGTTAACAACAGCTCGGATCCATGTCAC 1214  
QY 1402 -----CCCAACCCAGGAGAGGTGAACCCAG 1426  
DB 1215 TTGACGAGCAACAGGATACCATATGAAGCTGAGCCCAACCCAGGAGAGGTGAACCCAG 1274  
QY 1427 TTTTAAAGTTGCACTACAGCAGAGGAAGCAACTACAGCAACAATAAAGCTGAACCTTTA 1486  
DB 1275 TTTTAAAGTTGCACTACAGCAGAGGAAGCAACTACAGCAACAATAAAGCTGAACCTTTA 1334  
QY 1487 CAGATGAATGGAGCAGTATAGCATCAAGTCTTAGAGGAATTTGGAGCCATTTGCAATCTG 1546

1335 CAGATGAATGGAGCAGTATAGCATCAAGTTCTAGAGGAATTTGGGAGCCATTCGAATCTG 1394  
 1547 AGGGTCAGAGGAATCTTTTCGTCACAGAGTCTAGTCAACACCAAGAGGAGACAGTG 1606  
 1395 AGGGTCAGAGGAATCTTTTCGTCACAGAGTCTAGTCAACACCAAGAGGAGACAGTG 1454  
 1607 AAACAAAAGCTCTCTGAAGAATCATCAGAGGATGTGACAAAATATCAGGAGGAGTATCTG 1666  
 1455 AAACAAAAGCTCTCTGAAGAATCATCAGAGGATGTGACAAAATATCAGGAGGAGTATCTG 1514  
 1667 CAGAAAACCCAGTCTGAGAACCATATCAATATACCAATCAGATTAAGTTACAGGCCAAGC 1726  
 1515 CAGAAAACCCAGTCTGAGAACCATATCAATATACCAATCAGATTAAGTTACAGGCCAAGC 1574  
 1727 CATTTGATTCACCACTCAGAGAGAAAGAAATGACCTCAATCTTGATCGCTCTTGTGGGGTTC 1786  
 1575 CATTTGATTCACCACTCAGAGAGAAAGAAATGACCTCAATCTTGATCGCTCTTGTGGGGTTC 1634  
 1787 CAGAGAATCTGCTTCATCTGAAAAGCCAGAACCAAGAACCTTCAGATCAGCATAGCA 1846  
 1635 CAGAGAATCTGCTTCATCTGAAAAGCCAGAACCAAGAACCTTCAGATCAGCATAGCA 1694  
 1847 CTGAGAGTCTACCAATGAAAATTAACACCAATCTCTGAGCCTCAGTTCCAAACAGAGCCA 1906  
 1695 CTGAGAGTCTACCAATGAAAATTAACACCAATCTCTGAGCCTCAGTTCCAAACAGAGCCA 1754  
 1907 CTGGGCTTCAGCTCATGAAGAACATCCACAGGACCTCTGCTCTTCAGGACACAGATG 1966  
 1755 CTGGGCTTCAGCTCATGAAGAACATCCACAGGACCTCTGCTCTTCACACAGATG 1813  
 1967 ACAGTGATGATGACCAAGTCTCTGATCCAGTGCAGAGTATCAGCAGGACCTGGTGATA 2026  
 1814 ACAGTGATGATGACCAAGTCTCTGATCCAGTGCAGAGTATCAGCAGGACCTGGTGATA 1870  
 2027 GAGCCTCTGCTGTGCGCGTATTCAGAGTCTTCAGACGAGGAGAAAGAGAAAGAAA 2086  
 1871 GAGCCTCTGCTGTGCGCGTATTCAGAGTCTTCAGACGAGGAGAAAGAGAAAGAAA 1930  
 2087 TGAAGAATTTGGATACCTTGAACATTAAGAGCGCGTGTGATAAATGTTTATAAGGCC 2146  
 1931 TGAAGAATTTGGATACCTTGAACATTAAGAGCGCGTGTGATAAATGTTTATAAGGCC 1990  
 2147 ATCCAACTCCAGCAATGATAAAGAGCAATTTCTGGGGTCTAACTTTCTTAATGA 2206  
 1991 ATCCAACTCCAGCAATGATAAAGAGCAATTTCTGGGGTCTAACTTTCTTAATGA 2050  
 2207 GTGGTCTGACTGTGGCCACATTTTCATCTGGGATCGGCACACTGCTGAGCATTTGATGC 2266  
 2051 GTGGTCTGACTGTGGCCACATTTTCATCTGGGATCGGCACACTGCTGAGCATTTGATGC 2110  
 2267 TTCGGAAGCTGATTAATCATCTGTTAACTGCCTGAGCCACATCCGTTTGACCAATTT 2326  
 2111 TTCGGAAGCTGATTAATCATCTGTTAACTGCCTGAGCCACATCCGTTTGACCAATTT 2170  
 2327 TAGCCTCATCTGGCATATGATTAAGATCTGCTGACCATTTAGAAAGTCAAGGA 2386  
 2171 TAGCCTCATCTGGCATATGATTAAGATCTGCTGACCATTTAGAAAGTCAAGGA 2230  
 2387 TTTTAAACCGAAACCTTCTGATGAAGTTATAACTCGAAACGAACTCATGCTGGGAAGAA 2446  
 2231 TTTTAAACCGAAACCTTCTGATGAAGTTATAACTCGAAACGAACTCATGCTGGGAAGAA 2290  
 2447 CTAGAAACACCATTTACAGTTCCAGCCTCTTTCATCTTGAGGATGTTGGCTTCACATTAATC 2506  
 2291 CTAGAAACACCATTTACAGTTCCAGCCTCTTTCATCTTGAGGATGTTGGCTTCACATTAATC 2350  
 2507 ATATCCGAGCTCACCGGTTGGAGGTTGACAGATCAGAGGCTCTGCTCAAGAGAAATGAAA 2566  
 2351 ATATCCGAGCTCACCGGTTGGAGGTTGACAGATCAGAGGCTCTGCTCAAGAGAAATGAAA 2410  
 2567 ATGAGGATGAGGAA 2580  
 |||||

Db 2411 ATGAGGATGAGGAA 2424  
 RESULT 6  
 ABK35809  
 ID ABK35809 standard; cDNA; 1942 BP.  
 XX  
 AC ABK35809;  
 XX  
 DT 08-MAY-2002 (first entry)  
 XX  
 DE cDNA sequence #200 encoding novel human secreted protein.  
 XX  
 KW Human secreted protein; hyperproliferative disorder; autoimmune disorder;  
 KW immune deficiency disorder; blood disorder; inflammatory disorder;  
 KW infectious disorder; allergic condition; neurodegenerative disorder;  
 KW liver fibrosis; coagulation disorder; gene therapy; antimicrobial;  
 KW tumour; cancer; hepatotropic; immunosuppressive; antirheumatic; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200177289-A2.  
 XX  
 PD 18-OCT-2001.  
 XX  
 PF 29-MAR-2001; 2001WO-US10232.  
 XX  
 PR 06-APR-2000; 2000US-195605P.  
 XX  
 PA (GEM) GENETICS INST INC.  
 XX  
 PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;  
 PI Merberg D, Tracy M, Agostino MJ, Bowman MR, Spaulding V, Wong GG;  
 PI Clark HF, Fectel K, Howes SH, Resnick RJ, Gulukota K, Graham JR;  
 XX  
 DR WPI; 2002-179322/23.  
 XX  
 PT Six hundred and twenty three polynucleotides derived from a variety of  
 PT human tissue sources which encode secreted proteins, useful for  
 PT treating immune deficiencies and disorders such as autoimmune disorders  
 XX  
 PS Claim 1; Page 194-195; 393pp; English.  
 XX  
 CC The present invention relates to the isolation of novel cDNA sequences  
 CC which encode human secreted proteins. The cDNA sequences have been  
 CC derived from a variety of human tissues. The invention also provides  
 CC a method for producing proteins from these polynucleotide sequences.  
 CC The proteins are useful for identifying compounds that modulate their  
 CC activity and production. The sequences of the invention are  
 CC useful for treating diseases such as hyperproliferative disorders  
 CC (e.g. cancer), immune deficiency disorders (e.g. severe combined  
 CC immunodeficiency (SCID)), autoimmune disorders (e.g. multiple  
 CC sclerosis), blood disorders (e.g. thrombocytopaenia), inflammatory  
 CC disorders (e.g. arthritis), infectious disorders (e.g. hepatitis),  
 CC allergic conditions (e.g. asthma), neurodegenerative disorders (e.g.  
 CC Alzheimer's disease), liver fibrosis, coagulation disorders (e.g.  
 CC haemophilia), and tumours. The polynucleotide sequences of the  
 CC invention are also useful in gene therapy. ABK35610-ABK36232 represent  
 CC the cDNA sequences of the invention that encode for novel human  
 CC secreted proteins.  
 XX  
 SQ Sequence 1942 BP; 626 A; 383 C; 420 G; 513 T; 0 other;  
 Query Match 59.8%; Score 1542.4; DB 24; Length 1942;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 1543; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1037 CTGATATGTTATCAAGATGTTTGAAGAGCAGTGGTTCACAAAACCAATAGAGGAC 1096  
 Db 1 CTGATATGTTATCAAGATGTTTGAAGAGCAGTGGTTCACAAAACCAATAGAGGAC 60  
 QY 1097 GAGGAAGATCTCGACCCAGAGGTGGAAACCAATCAGATATTTCAACTCTTCCTACGG 1156  
 |||||



XX	SQ	Sequence	1952 BP;	642 A;	392 C;	418 G;	500 T;	0 other;	
		Query Match	56.1%	Score	1446.4;	DB	24;	Length	1952;
		Best Local Similarity	96.1%	Pred. No.	0;				
		Matches	1517;	Conservative	0;	Mismatches	1;	Indels	60;
									Gaps
QY	1063	GAACCAAGTCAGGTTGCACAAAGCAATAGAGGACGAGGAAGATCTCGACCCAGAGCTGGA	1122						
DB	1	GAACCAAGTCAGGTTGCACAAAGCAATAGAGGACGAGGAAGATCTCGACCCAGAGCTGGA	60						
QY	1123	ACAAAGTCAATCAGATATTTCAACTTCTTCCACGGTCCCATCAAGTCCTGATTTGGAAGTG	1182						
DB	61	ACAAGTCAATCAGATATTTCAACTTCTTCCACGGTCCCATCAAGTCCTGATTTGGAAGTG	120						
QY	1183	AGTGAAGCTGCAATGGGAATGATATCTCCAGCTGAAACAATTTCTTCAGCCTCTTACATCC	1242						
DB	121	AGTGAAGCTGCAATGGGAATGATATCTCCAGCTGAAACAATTTCTTCAGCCTCTTACATCC	180						
QY	1243	TCTCAATATGTCAGCTCAGGCTCATTCGACATCATCTCCACAGAAAGCCCTCATTTCTACT	1302						
DB	181	TCTCAATATGTCAGCTCAGGCTCATTCGACATCATCTCCACAGAAAGCCCTCATTTCTACT	240						
QY	1303	CGTTTGCCTATCTTCCAGACAGTGAACAAAGGCAGTCTGTTGAGGCATCTGGACACCCAC	1362						
DB	241	CGTTTGCCTATCTTCCAGACAGTGAACAAAGGCAGTCTGTTGAGGCATCTGGACACCCAC	300						
QY	1363	ACACATCATCAGTCT - - - - -	1377						
DB	301	ACACATCATCAGTCTGATCTTCTTCTGTGGTTAAACAACAGCTCGGATCCATGTCA	360						
QY	1378	- - - - - GATAACAATATGAAAGCTGAGCGCCCAACACAGGACAGGTGAA	1422						
DB	361	CTTGACGAGCAACAGGATTAACAATATGAAAGCTGAGCGCCCAACACAGGACAGGTGAA	420						
QY	1423	CCAGTTTAAAGTTTGCATCTACGACACAGAGGAACAACATACAAGCACAAATAAACTGAAAC	1482						
DB	421	CCAGTTTAAAGTTTGCATCTACGACACAGAGGAACAACATACAAGCACAAATAAACTGAAAC	480						
QY	1483	TTTACAGATGAATGGACGAGTATAGCATCAAGTTCTAGAGGAATTTGGAGCCATTTGCAAA	1542						
DB	481	TTTACAGATGAATGGAGCAGTATAGCATCAAGTTCTAGAGGAATTTGGAGCCATTTGCAAA	540						
QY	1543	CTTGAGGGTCAGAGGAATCTTTTGTCACACAGAGCTCAGTGCACACACAGGAGAGAC	1602						
DB	541	CTTGAGGGTCAGAGGAATCTTTTGTCACACAGAGCTCAGTGCACACACAGGAGAGAC	600						
QY	1603	AGTGAACAACAAAGCTCCTGAAGAATCATCAGAGGATGTGACAAAAATATCAGGAAGAGTA	1662						
DB	601	AGTGAACAACAAAGCTCCTGAAGAATCATCAGAGGATGTGACAAAAATATCAGGAAGAGTA	660						
QY	1663	CTTGCGAAAAACCCAGTTGAGAACCATATCAATATAACACAATCAGATTAAGTTTCAGGCC	1722						
DB	661	CTTGCGAAAAACCCAGTTGAGAACCATATCAATATAACACAATCAGATTAAGTTTCAGGCC	720						
QY	1723	AAGCCATTGATTTCCAACTCAGGAGAAAGAAATGACCTCAATCTTCATCGCTCTTGTGGG	1782						
DB	721	AAGCCATTGATTTCCAACTCAGGAGAAAGAAATGACCTCAATCTTCATCGCTCTTGTGGG	780						
QY	1783	GTTCCAGAAAGAAATCTGCTTCATCTGAAAAAGCCAAAGGAACCCAGAAATCTCAGATCAGACT	1842						
DB	781	GTTCCAGAAAGAAATCTGCTTCATCTGAAAAAGCCAAAGGAACCCAGAAATCTCAGATCAGACT	840						
QY	1843	AGCACTGAGAGTGTACCAATGAATAATAACACCAATCCTGAGCCTCAGTTTCCAAAACAGAA	1902						
DB	841	AGCACTGAGAGTGTACCAATGAATAATAACACCAATCCTGAGCCTCAGTTTCCAAAACAGAA	900						
QY	1903	GCACCTGGCCCTCAGCTCATGAAGAAACATCCACACAGGACTCTGCTCTTCAGGACACA	1962						
DB	901	GCACCTGGCCCTCAGCTCATGAAGAAACATCCACACAGGACTCTGCTCTTCAGGACACA	960						
QY	1963	GATCACAGTGTATGATGACCCAGTCTCTGATCCAGGTCGAAGGATATCAGCAGGACCTGGT	2022						

Db	961	GATCACAGCTGATGATGACCCAGTCTCTGATCCAGGTGCAAGGTATCGACGAGGACCTGGT	1020
Qy	2023	CATACAGCCTCTGCTGTTGCCGATTTCAGGAGTCTTTCAGACGAGAGAAAAGAAAGAAA	2082
Db	1021	GATAGACGCTCTGCTGTTGCCGATTTCAGGAGTCTTTCAGACGAGAGAAAAGAAAGAAA	1080
Qy	2083	GAATATGGAAGAATTTGGATACCTTTGAACATTTAGAAGCGCGCTAGTAAATGTTTTATAAA	2142
Db	1081	GAATATGGAAGAATTTGGATACCTTTGAACATTTAGAAGCGCGCTAGTAAATGTTTTATAAA	1140
Qy	2143	GGCCATCGCACTCCAGACGATGATAAAGAAAGCAATTTCTGGGGTCTAACTTTGTGA	2202
Db	1141	GGCCATCGCACTCCAGACGATGATAAAGAAAGCAATTTCTGGGGTCTAACTTTGTGA	1200
Qy	2203	ATGAGTGGTTTCTGACTGTGGCCACATTTTTCATCTGGGATCGGCACACTGCTGAGCATTTG	2262
Db	1201	ATGAGTGGTTTCTGACTGTGGCCACATTTTTCATCTGGGATCGGCACACTGCTGAGCATTTG	1260
Qy	2263	ATGCTTCTGGAAGCTGATTAATCATGTGGTAAACTGCTGCAGCCACATCCGTTTGACCCA	2322
Db	1261	ATGCTTCTGGAAGCTGATTAATCATGTGGTAAACTGCTGCAGCCACATCCGTTTGACCCA	1320
Qy	2323	ATTTTAGCCTCATCTGGCATAGATTATGACATAAAGATCTGTCACCATTAGAAGAGTCA	2382
Db	1321	ATTTTAGCCTCATCTGGCATAGATTATGACATAAAGATCTGTCACCATTAGAAGAGTCA	1380
Qy	2383	AGGATTTTTAACCGAAAACCTTGTCTGATGAGTTTAACTCGAAACGAACATCATGCTGGAA	2442
Db	1381	AGGATTTTTAACCGAAAACCTTGTCTGATGAGTTTAACTCGAAACGAACATCATGCTGGAA	1440
Qy	2443	GAACATAGAAACACCACTTACAGTTCCAGCCTCTTTCATGTTGAGGATGTTGGCTTCACTT	2502
Db	1441	GAACATAGAAACACCACTTACAGTTCCAGCCTCTTTCATGTTGAGGATGTTGGCTTCACTT	1500
Qy	2503	AATCATATCCGAGCTGACCGGTTGGAGGTTGACAGATCAGAAGCCTCTGGTCAAGAGAAT	2562
Db	1501	AATCATATCCGAGCTGACCGGTTGGAGGTTGACAGATCAGAAGCCTCTGGTCAAGAGAAT	1560
Qy	2563	GAAAATCAGGATGAGGAA 2580	
Db	1561	GAAAATCAGGATGAGGAA 1578	
RESULT 8			
AAH13776			
ID	AAH13776 standard; cdna; 1726 BP.		
XX	AAH13776;		
XX	26-JUN-2001 (first entry)		
XX	Human cdna sequence SEQ ID NO:10702.		
XX	Human; primer; detection; diagnosis; antisense therapy; gene therapy;		
XX	Homo sapiens.		
XX	EP1074617-A2.		
XX	07-FEB-2001.		
XX	28-JUL-2000; 2000EP-0116126.		
XX	29-JUL-1999; 99JP-0248036.		
XX	27-AUG-1999; 99JP-0300253.		
XX	11-JAN-2000; 2000JP-0118776.		
XX	02-MAY-2000; 2000JP-0183767.		
XX	09-JUN-2000; 2000JP-0241899.		
XX	(HELI-) HELIX RES INST.		
XX	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;		
XX	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T		





[illegible]

Query Match 30.9%; Score 798.4; DB 22; Length 1012;



XX Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.  
 XX Homo sapiens.  
 OS WO200166719-A1.  
 PN 13-SEP-2001.  
 PD  
 XX  
 XX 02-MAR-2001; 2001WO-JP01629.  
 XX 07-MAR-2000; 2000JP-0159195.  
 XX (CHIB-) CHIBA PREFECTURE.  
 XX (HISM) HISAMITSU PHARM CO LTD.  
 XX Nakagawara A;  
 PI WPI; 2001-565584/63.  
 DR  
 XX Nucleic acids originating in gene expressed in human neuroblastoma,  
 XX useful as probe or primer in diagnosing prognosis of human  
 XX neuroblastoma, malignancy and susceptibility indicator or tumour marker  
 XX for anti-cancer agents  
 XX  
 PS Claim 1; Page 2858; 2979pp; Japanese.  
 XX  
 CC The invention relates to novel genes (AA193926-AA197963) expressed in  
 CC human neuroblastoma. The nucleic acids are applicable as a probe or  
 CC primer in diagnosing the prognosis of human neuroblastoma, malignancy and  
 CC susceptibility indicators or tumour markers for anti-cancer agents. The  
 CC gene information for diagnosing prognosis is related to factors similar  
 CC to that for N-myc and TrkA genes.  
 XX  
 SQ Sequence 728 BP; 231 A; 159 C; 169 G; 167 T; 2 other;  
 Query Match 26.0%; Score 671.4; DB 22; Length 728;  
 Best Local Similarity 99.6%; Pred. No. 7.9e-182;  
 Matches 694; Conservative 0; Mismatches 1; Indels 2; Gaps 2;  
 QY 830 TATATCTTTTGGCCGAAAGATGATACGACGAGAGACTTAAACCTCTCTCGCGAAG 889  
 DB 32 TATATCTTTTGGCCGAAAGATGATACGACGAGAGACTTAAACCTCTCTCGCGAAG 91  
 QY 890 AGAGAAGAGAGAGTTGCGACACACACAGTTAAGCGTTTGAGACTTCGTGGTATGGT 949  
 DB 92 AGAGAAGAGAGAGTTGCGACACACACAGTTAAGCGTTTGAGACTTCGTGGTATGGT 151  
 QY 950 CAGATCTGGACCCAGAGCAAGCGCGGAGTGACGAGAGAGATGGAGAGCAGAGTC 1009  
 DB 152 CAGATCTGGACCCAGAGCAAGCGCGGAGTGACGAGAGAGATGGAGAGCAGAGTC 211  
 QY 1010 CCAATGTGTCTATTGATGACAGAGATCTCTGATATGTTATCAAGATGGTTGAAGAGCAA 1069  
 DB 212 CCAATGTGTCTATTGATGACAGAGATCTCTGATATGTTATCAAGATGGTTGAAGAGCAA 271  
 QY 1070 GTGAGTTGCCAACAAGCAATAGAGGAGGAGAGATCTCGACCCAGAGTGGAACAAGTC 1129  
 DB 272 GTGAGTTGCCAACAAGCAATAGAGGAGGAGAGATCTCGACCCAGAGTGGAACAAGTC 331  
 QY 1130 AATCAGATATTTCAACTCTTCTACGTCGCCATCAAGTCCTGATTTGGAGTGAGTGA 1189  
 DB 332 AATCAGATATTTCAACTCTTCTACGTCGCCATCAAGTCCTGATTTGGAGTGAGTGA 391  
 QY 1190 CTGCAATGGAAGTAGATCTCCAGCTGACAAATTTCTTCAGCTTCTACATCTCTACAA 1249  
 DB 392 CTGCAATGGAAGTAGATCTCCAGCTGACAAATTTCTTCAGCTTCTACATCTCTACAA 451  
 QY 1250 TGTACGCTCAGGCTCATTCGACATCATCTCCACAGAAGCCCTCATTTACTCTTTGC 1309  
 DB 452 TGTACGCTCAGGCTCATTCGACATCATCTCCACAGAAGCCCTCATTTACTCTTTGC 511  
 QY 1310 TATCTTCTCCAGACAGTGAACAAAGGCGAGTCTGTTGAGGCATCTGGACACACACATC 1369

DB 512 TATCTTCTCAGACAGTGAACAAAGCGAGTCTGTTGAGGCATCTGACACCCACACATC 571  
 QY 1370 ATCAGTCTGATAACAATAATGAAAGCTGAGCCCCAAACAGGAGCAGGTGAACACAGTTT 1429  
 DB 572 ATCAGTCTGATAACAATAATGAGAGCTGAGCCCCAAACAGGAGCAGGTGAACACAGTTT 631  
 QY 1430 TAAGTTTGCACTACAGCAGAGGAGAACTACAAGCACAATAAATCACTGAC-TTTACA 1488  
 DB 632 TAAGTTTGCACTACAGCAGAGGAGAACTACAAGCACAATAAATCACTGAC-TTTACA 691  
 QY 1489 GATGAAT-GGAGCAGTATAGCATCAAGTTCTAGAGGA 1524  
 DB 692 GATGAATGGAGCAGTATAGCATCAAGTTCTAGAGGA 728  
 RESULT 11  
 AAA44677  
 ID AAA44677 standard; cDNA; 635 BP.  
 XX  
 AC AAA44677;  
 XX  
 DT 21-AUG-2000 (first entry)  
 XX  
 DE Human secreted expressed sequence tag SEQ ID NO:1252.  
 KW Human; mouse; chicken; rat; secreted expressed sequence tag; sEST;  
 KW expressed sequence tag; EST; probe; chemotactic; proliferative;  
 KW immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;  
 KW thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;  
 KW antiviral; antidiabetic; antiasthmatic; vulnerary; antiparkinsonian;  
 KW antitumor; osteopathic; neuroprotective; nontropic; antipsoriatic;  
 KW cerebroprotective; anticonvulsant; antidepressant; gene therapy;  
 KW vaccine; autoimmune disorder; multiple sclerosis; allergic condition;  
 KW insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;  
 KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;  
 KW central nervous system disorder; Alzheimer's disease; stroke;  
 KW Parkinson's disease; Huntington's disease; coagulation disorder;  
 KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;  
 KW tumour; infection; depression; psoriasis; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200021991-A1.  
 XX  
 PD 20-APR-2000.  
 XX  
 PF 15-OCT-1999; 99WO-US24206.  
 XX  
 PR 15-OCT-1998; 98US-0104436.  
 XX  
 XX (GEMY) GENETICS INST INC.  
 XX  
 XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;  
 PI Merberg D, Treacy M, Bowman MR;  
 XX  
 XX WPI; 2000-317938/27.  
 XX  
 PT Isolated polynucleotides, and encoded proteins, comprising secreted  
 PT expressed sequence tags (sESTs), useful for treating various disorders  
 PT such as autoimmune, infectious, and central nervous system disorders -  
 XX  
 PS Claim 1; Page 527; 803pp; English.  
 XX  
 CC AAA43426 to AAA45925 represent specifically claimed secreted expressed  
 CC sequence tags (sESTs), isolated from human, mouse, chicken and rat  
 CC tissue sources. The sESTs can have a range of activities depending on  
 CC the tissues they were isolated from. The activities include:  
 CC chemotactic; proliferative; immunomodulatory; haematopoietic;  
 CC chemokinetic; analgesic; haemostatic; thrombolytic; antidiabetic;  
 CC cytostatic; antibacterial; antifungal; antitumor; antipsoriatic;  
 CC antiasthmatic; vulnerary; antitumor; osteopathic; neuroprotective;  
 CC nontropic; antiparkinsonian; antipsoriatic; cerebroprotective;

QY	1442	ACAGCACAAGGAACAACACTCAAGCAACAATAAAAACCTGAACTTTACAGATGAATGGAGCA	1501
Db	240	AAAGATCATATTGATTAACTTCTTGGTGAATCACTTATCTTGTGTTTGTATGATAGAGCA	299
QY	1502	GTATAGCATCAAGTTCTAGAGAAATGGAGCCATTGCCAAATCTCAGGGTTCAGGAGGAAT	1561
Db	300	GTATAGCATCAAGTTCTAGAGAAATGGAGCCATTGCCAAATCTCAGGGTTCAGGAGGAAT	359
QY	1562	CTTTTCGTCCACAGAGCTCAGTGCACACCACAGAGGACAGTGAACAAAAGTCTCTG	1621
Db	360	CTTTTCGTCCACAGAGCTCAGTGCACACCACAGAGGACAGTGAACAAAAGTCTCTG	419



RESULT 16  
AAK41053  
ID AAK41053 standard; DNA; 332 BP.  
XX  
AC AC  
XX AAK41053;  
XX

## RESULT 17

1700	Y	CACAATCAGATAAGCTTCACAGCCAGCCATTGGATCCAACTCAGGAGAAAGAAATGACC	1759
	bb		
1	Y	CACAATCAGATAAGTTACAGACCCAGCCATTGGATTCCAACTCAGGAGAAAGAAATGACC	60
	bb		
1760	Y	TCATCTTTGATCGCTCTGTGGGTTCCAGAGAAATCTGCTTCATCTGAAAAAGCCAAAG	1819
	bb		
61	Y	TCATCTTTGATCGCTCTGTGGGTTCCAGAGAAATCTGCTTCATCTGAAAAAGCCAAAG	120
	Y		
1820	Y	AACCAGAAACTTCAGATCAGACTAGACACTGAGAGTGTACCAATGAAATACACCAATC	1879
	b		
121	Y	AACCAGAAACTTCAGATCAGACTAGACACTGAGAGTGTACCAATGAAATACACCAATC	180
	Y		
1880	Y	CTGAGCCTCAGTTCCAAACAGAACCCACTGGGCCCTTCAGCTCATGAGAAACATCCACCA	1939
	b		
181	Y	CTGAGCCTCAGTTCCAAACAGAACCCACTGGGCCCTTCAGCTCATGAGAAACATCCACCA	240
	Y		
1940	Y	GGGAGCTCTGCTTTCAGGACACAGATGACAGTGTATGATGACCCAGTCCCTGATCCAGGTG	1999

db	61	TCATCTTGATCGCTCTTGTGGGGTTCGAGAAATCTGCTTCATCTGAAAAGCCAAG	120
Yy	1820	AACAGAAACTTCAGATCAGACTAGCAGCTGAGAGTGCTACCAATGAATAATACACCAATC	1879
db	121	AACAGAAACTTCAGATCAGACTAGCAGCTGAGAGTGCTACCAATGAATAATACACCAATC	180
Yy	1880	CTGAGCCTCAGTTCCTCAACAGAGCCACTGGGCCCTTCAGCTCATGAGAACAATCCACCA	1939
db	181	CTGAGCCTCAGTTCCTCAACAGAGCCACTGGGCCCTTCAGCTCATGAGAACAATCCACCA	240
Yy	1940	GGGACTCTGCTCTTCAGACACAGATGACACTGATGATGACCCAGTCCTGATCCCAGGTG	1999
db	241	GGGACTCTGCTCTTCAGACACAGATGACACTGATGATGACCCAGTCCTGATCCCAGGTG	300
Yy	2000	CAAGGTATCGACGAGCACTGGTGATAG	2027
db	301	CAAGGTATCGACGAGCACTGGTGATAG	328

RESULT 19	
ABSI5038	
ID ABSI5038 standard; DNA; 332 BP.	
XX	
XX	
AC	
ABS15038;	
19-AUG-2002 (first entry)	
DT	
DT	
DE	
DE	
KW	Human genome-derived single exon probe ORF from lung SEQ ID No 15029.
KW	Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
KW	chronic obstructive pulmonary disease; interstitial lung disease;
KW	familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW	tubercous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW	Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemodierosis;
KW	pulmonary histiocytosis; lymphangioloelomyomatosis; Karagener syndrome;
KW	pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW	primary ciliary dyskinesia; pulmonary hypertension;
KW	reading frame; ORF.

Homo sapiens.  
 WO200186003-A2.  
 15-NOV-2001.  
 30-JAN-2001; 2001WO-US00665.  
 04-FEB-2000; 2000US-180312P.  
 26-MAY-2000; 2000US-207456P.  
 30-JUN-2000; 2000US-0608408.  
 03-AUG-2000; 2000US-0632366.  
 21-SEP-2000; 2000US-234687P.  
 27-SEP-2000; 2000US-236359P.  
 04-OCT-2000; 2000GB-0024263.  
 (MOLE-) MOLECULAR DYNAMICS INC.  
 Penn SG, Hanzel DK, Chen W, Rank DR;  
 wptl: 2002-114183/15.

spatially-addressable set of single exon nucleic acid probes, used to measure gene expression in human lung samples - English.

Claim 4; SEQ ID NO 10057, 10058, 10059, 10060, 10061, 10062, 10063, 10064, 10065, 10066, 10067, 10068, 10069, 10070, 10071, 10072, 10073, 10074, 10075, 10076, 10077, 10078, 10079, 10080, 10081, 10082, 10083, 10084, 10085, 10086, 10087, 10088, 10089, 10090, 10091, 10092, 10093, 10094, 10095, 10096, 10097, 10098, 10099, 10100, 10101, 10102, 10103, 10104, 10105, 10106, 10107, 10108, 10109, 10110, 10111, 10112, 10113, 10114, 10115, 10116, 10117, 10118, 10119, 10120, 10121, 10122, 10123, 10124, 10125, 10126, 10127, 10128, 10129, 10130, 10131, 10132, 10133, 10134, 10135, 10136, 10137, 10138, 10139, 10140, 10141, 10142, 10143, 10144, 10145, 10146, 10147, 10148, 10149, 10150, 10151, 10152, 10153, 10154, 10155, 10156, 10157, 10158, 10159, 10160, 10161, 10162, 10163, 10164, 10165, 10166, 10167, 10168, 10169, 10170, 10171, 10172, 10173, 10174, 10175, 10176, 10177, 10178, 10179, 10180, 10181, 10182, 10183, 10184, 10185, 10186, 10187, 10188, 10189, 10190, 10191, 10192, 10193, 10194, 10195, 10196, 10197, 10198, 10199, 10200, 10201, 10202, 10203, 10204, 10205, 10206, 10207, 10208, 10209, 10210, 10211, 10212, 10213, 10214, 10215, 10216, 10217, 10218, 10219, 10220, 10221, 10222, 10223, 10224, 10225, 10226, 10227, 10228, 10229, 10230, 10231, 10232, 10233, 10234, 10235, 10236, 10237, 10238, 10239, 10240, 10241, 10242, 10243, 10244, 10245, 10246, 10247, 10248, 10249, 10250, 10251, 10252, 10253, 10254, 10255, 10256, 10257, 10258, 10259, 10260, 10261, 10262, 10263, 10264, 10265, 10266, 10267, 10268, 10269, 10270, 10271, 10272, 10273, 10274, 10275, 10276, 10277, 10278, 10279, 10280, 10281, 10282, 10283, 10284, 10285, 10286, 10287, 10288, 10289, 10290, 10291, 10292, 10293, 10294, 10295, 10296, 10297, 10298, 10299, 10300, 10301, 10302, 10303, 10304, 10305, 10306, 10307, 10308, 10309, 10310, 10311, 10312, 10313, 10314, 10315, 10316, 10317, 10318, 10319, 10320, 10321, 10322, 10323, 10324, 10325, 10326, 10327, 10328, 10329, 10330, 10331, 10332, 10333, 10334, 10335, 10336, 10337, 10338, 10339, 10340, 10341, 10342, 10343, 10344, 10345, 10346, 10347, 10348, 10349, 10350, 10351, 10352, 10353, 10354, 10355, 10356, 10357, 10358, 10359, 10360, 10361, 10362, 10363, 10364, 10365, 10366, 10367, 10368, 10369, 10370, 10371, 10372, 10373, 10374, 10375, 10376, 10377, 10378, 10379, 10380, 10381, 10382, 10383, 10384, 10385, 10386, 10387, 10388, 10389, 10390, 10391, 10392, 10393, 10394, 10395, 10396, 10397, 10398, 10399, 10400, 10401, 10402, 10403, 10404, 10405, 10406, 10407, 10408, 10409, 10410, 10411, 10412, 10413, 10414, 10415, 10416, 10417, 10418, 10419, 10420, 10421, 10422, 10423, 10424, 10425, 10426, 10427, 10428, 10429, 10430, 10431, 10432, 10433, 10434, 10435, 10436, 10437, 10438, 10439, 10440, 10441, 10442, 10443, 10444, 10445, 10446, 10447, 10448, 10449, 10450, 10451, 10452, 10453, 10454, 10455, 10456, 10457, 10458, 10459, 10460, 10461, 10462, 10463, 10464, 10465, 10466, 10467, 10468, 10469, 10470, 10471, 10472, 10473, 10474, 10475, 10476, 10477, 10478, 10479, 10480, 10481, 10482, 10483, 10484, 10485, 10486, 10487, 10488, 10489, 10490, 10491, 10492, 10493, 10494, 10495, 10496, 10497, 10498, 10499, 10500, 10501, 10502, 10503, 10504, 10505, 10506, 10507, 10508, 10509, 10510, 10511, 10512, 10513, 10514, 10515, 10516, 10517, 10518, 10519, 10520, 10521, 10522, 10523, 10524, 10525, 10526, 10527, 10528, 10529, 10530, 10531, 10532, 10533, 10534, 10535, 10536, 10537, 10538, 10539, 10540, 10541, 10542, 10543, 10544, 10545, 10546, 10547, 10548, 10549, 10550, 10551, 10552, 10553, 10554, 10555, 10556, 10557, 10558, 10559, 10560, 10561, 10562, 10563, 10564, 10565, 10566, 10567, 10568, 10569, 10570, 10571, 10572, 10573, 10574, 10575, 10576, 10577, 10578, 10579, 10580, 10581, 10582, 10583, 10584, 10585, 10586, 10587, 10588, 10589, 10590, 10591, 10592, 10593, 10594, 10595, 10596, 10597, 10598, 10599, 10600, 10601, 10602, 10603, 10604, 10605, 10606, 10607, 10608, 10609, 10610, 10611, 10612, 10613, 10614, 10615, 10616, 10617, 10618, 10619, 10620, 10621, 10622, 10623, 10624, 10625, 10626, 10627, 10628, 10629, 10630, 10631, 10632, 10633, 10634, 10635, 10636, 10637, 10638, 10639, 10640

dbb 241 GGGACTGCTCTTCAGGACACAGATGACAGTGATGATGATGCCCGTCTGTGATCCCGGTG 300

QY 2000 CAAGGTATCGAGCAGGACCTGGTGATG 2027

2001 CGCGATCCGACGAGCACTGGTGATG 328

RESULT 18	
AA107503	
ID	AA107503 standard; DNA; 332 BP.
XX	
XX	AA107503;
XX	
XX	
XX	09-OCT-2001 (first entry)
XX	
XX	Probe #7494 used to measure gene expression in human breast sample.
XX	
XX	probe: human; breast disease; breast cancer; development disorder; ss;
XX	en-carcinoma tumour.

X	Homo sapiens.	
S		
X	WO200157270-A2.	
X		
N		
N		
X		
X		
D	09-AUG-2001.	
X		
X		
X	29-JAN-2001; 2001WO-US000661.	
X		
X	04-FEB-2000; 2000US-0180312.	
R	26-MAY-2000; 2000US-0207456.	
R	30-JUN-2000; 2000US-0608408.	
R	03-AUG-2000; 2000US-0632366.	
R	21-SEP-2000; 2000US-0234687.	
R	27-SEP-2000; 2000US-0236359.	
R	04-OCT-2000; 2000GB-0024263.	
XX		
XX	(MOLE-) MOLECULAR DYNAMICS INC.	
XX		
XX	Penn SG, Hanzel DK, Chen W, Rank DR;	
PI		
XX		
XX	WPI; 2001-476286/51.	
XX		
XX	Novel single exon nucleic acid probe used to measuring gene expression	
PT		
PT	in a human breast -	
XX		

Claim 25: SEQ ID NO 7434, 524557-524561

xx The present invention relates to novel single exon nucleic acid probes.  
xx The present sequence is one such probe. The probes are useful for  
cc measuring human gene expression in a human breast sample, where the probe  
cc hybridises at high stringency to a nucleic acid expressed in the human  
cc breast. The probes are useful for predicting, diagnosing, grading,  
cc staging, monitoring and prognosing diseases of the human breast,  
cc particularly those diseases with polygenic aetiology. The diseases  
cc include: breast cancer, disorders of development, inflammatory diseases  
cc of the breast, fibrocystic changes, proliferative breast disease and  
cc non-carcinoma tumours.  
cc Note: The sequence data for this patent did not form part of the printed  
cc specification, but was obtained in electronic format directly from WIPO  
cc database sequences.

CC at ftp.wipo.int/pub/published\_publications  
XX Sequence 332 BP; 107 A; 84 C; 74 G; 67 T; 0 other;  
SQ  
Query Match 12.7%; Score 328; DB 22; Length 332;  
Best Local Similarity 100.0%; Pred. No. 1.5e-83;  
Gaps 0; Indels 0;

[illegible]

CC probes: the novel set of probes which hybridise at high stringency to a  
 CC nucleic acid expressed in the human lung; measuring gene expression in a  
 CC sample derived from human lung, comprising (a) contacting the array with  
 CC a collection of detectably labeled nucleic acids derived from human lung  
 CC mRNA, and (b) measuring the label detectably bound to each probe of  
 CC the array; identifying exons in a eukaryotic genome, comprising  
 CC (a) algorithmically predicting at least one exon from genomic sequences  
 CC of the eukaryote; and (b) detecting specific hybridisation of detectably  
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
 CC having a fragment identical to the predicted exon, the probe is included  
 CC in the above mentioned microarray; assigning exons to a single gene,  
 CC comprising (a) identifying exons from genomic sequence by the method  
 CC above and (b) measuring the expression of each of the exons in several  
 CC tissues and/or cell types using hybridisation to a single exon  
 CC microarrays having a probe with the exon, where a common pattern of  
 CC expression of the exons in the tissues and/or cell types indicates that  
 CC the exons should be assigned to a single gene; a peptide comprising one  
 CC of 12011 sequences, mentioned in the specification, or encoded by the  
 CC probes/open reading frames (ORF). The probes are used for gene  
 CC expression analysis, and for identifying exons in a gene, particularly  
 CC using human lung derived mRNA and for the study of lung diseases  
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease  
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary  
 CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,  
 CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary  
 CC haemolysis, pulmonary histiocytosis, lymphangioleiomyomatosis,  
 CC pulmonary alveolar proteinosis, Karagen syndrome, fibrocystic  
 CC and hyaline membrane disease. The present sequence is a single exon  
 CC probe open reading frame of the invention.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIP0 at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 332 BP; 107 A; 84 C; 74 G; 67 T; 0 other;

Query Match	12.7%;	Score 328;	DB 24;	Length 332;
Best Local Similarity	100.0%;	Pred. No. 1.5e-83;		
Matches 328;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 1700 CACATCAGATAAGTTTCACAGCCAGCCATTGGATTCCCACTCAGGAGAAATGACC 1759  
 Db 1 CACATCAGATAAGTTTCACAGCCAGCCATTGGATTCCCACTCAGGAGAAATGACC 60  
 QY 1760 TCAATCTGTGTCGTCTGTGGGTTCCAGAGAAATCTGCTTCATCTGAAAAGCCAAAG 1819  
 Db 61 TCAATCTGTGTCGTCTGTGGGTTCCAGAGAAATCTGCTTCATCTGAAAAGCCAAAG 120  
 QY 1820 AACAGAAACTTCAGATCAGACTAGCAGTACAGTGTGCTTCAATGAAATTAACCAATC 1879  
 Db 121 AACAGAAACTTCAGATCAGACTAGCAGTACAGTGTGCTTCAATGAAATTAACCAATC 180  
 QY 1880 CTGAGCCTCAGTCCCAACAGAGCCACTGGGCTTCAGCTCATGAGAGAAATCCACCA 1939  
 Db 181 CTGAGCCTCAGTCCCAACAGAGCCACTGGGCTTCAGCTCATGAGAGAAATCCACCA 240  
 QY 1940 GGGACTCTGCTCTTCAGGACACAGATGACAGTGTGATGATGATGATGATGATGATG 1999  
 Db 241 GGGACTCTGCTCTTCAGGACACAGATGACAGTGTGATGATGATGATGATGATGATG 300  
 QY 2000 CAAGGTATCCAGACGACCTGGTGATG 2027  
 Db 301 CAAGGTATCCAGACGACCTGGTGATG 328

RESULT 20  
 AA197812/c  
 ID AA197812 standard; cDNA; 710 BP.  
 XX AA197812;  
 AC AA197812;  
 XX DT 13-NOV-2001 (first entry)

XX Human neuroblastoma expressed polynucleotide SEQ ID NO 3887.  
 DE Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.  
 KW Homo sapiens.  
 XX WO200166719-A1.  
 PN 13-SEP-2001.  
 PD 02-MAR-2001; 2001WO-JP01629.  
 PE 07-MAR-2000; 2000JP-0159195.  
 PR (CHIB-) CHIBA PREFECTURE.  
 XX (HISM) HISAMITSU PHARM CO LTD.  
 XX Nakagawara A;  
 DR WPI; 2001-565584/63.  
 XX Nucleic acids originating in gene expressed in human neuroblastoma,  
 PT useful as probe or primer in diagnosing prognosis of human  
 PT neuroblastoma, malignancy and susceptibility indicator or tumour marker  
 XX for anti-cancer agents -  
 PS Claim 1; Page 2859; 2979pp; Japanese.

CC The invention relates to novel genes (AA193926-AA197963) expressed in  
 CC human neuroblastoma. The nucleic acids are applicable as a probe or  
 CC primer in diagnosing the prognosis of human neuroblastoma, malignancy and  
 CC susceptibility indicators or tumour markers for anti-cancer agents. The  
 CC gene information for diagnosing prognosis is related to factors similar  
 CC to that for N-myc and TrkA genes.

Sequence 710 BP; 218 A; 139 C; 109 G; 240 T; 4 other;

Query Match	11.2%;	Score 287.8;	DB 22;	Length 710;
Best Local Similarity	98.4%;	Pred. No. 7.8e-72;		
Matches 300;	Conservative 0;	Mismatches 4;	Indels 1;	Gaps 1;

QY 2277 TGATTAATCATGTGGTAAACTGCCTGCAGCC-ACATCCGTTTGACCCCAATTTAGCCTCAT 2335  
 Db 699 TAATAATCATGTGGTAAACTGCCTGCAGCCACCAACATCCGTTGACCCCAATTTAGCCTCAT 640  
 QY 2336 CTGGCATAGATTATGACATATAGATCTGGTCAACATTAGAAGAGTCAAGGATTTTAAAC 2395  
 Db 639 CTGGCATAGATTATGACATATAGATCTGGTCAACATTAGAAGAGTCAAGGATTTTAAAC 580  
 QY 2396 GAAAACCTTGCTGATGAAGTTATACTCGAAACGAACTCATCTGGAAGAACTAGAAACA 2455  
 Db 579 GAAAACCTTGCTGATGAAGTTATACTCGAAACGAACTCATCTGGAAGAACTAGAAACA 520  
 QY 2456 CCATTACAGTTCCAGCCTCTTTCATGTTGAGGATGTTGGCTTCACTTAATCATATCCGAG 2515  
 Db 519 CCATTACAGTTCCAGCCTCTTTCATGTTGAGGATGTTGGCTTCACTTAATCATATCCGAG 460  
 QY 2516 CTGACCGGTTGGAGGTTACAGATCAGAGGCTCTGCTCAAGAGAAATGAAATGAGGATG 2575  
 Db 459 CTGACCGGTTGGAGGTTACAGATCAGAGGCTCTGCTCAAGAGAAATGAAATGAGGATG 400  
 QY 2576 AGGAA 2580  
 Db 399 AGGAA 395

RESULT 21  
 ABA46979  
 ID ABA46979 standard; DNA; 261 BP.  
 XX ABA46979;  
 AC ABA46979;  
 XX DT 13-NOV-2001 (first entry)

|||||

Db 181 CTACTCTTTGCTATCTTCCAGACAGTGACAAAGGCGAGTCTGTGGCATCTGGAC 240

Qy 1358 ACCACACACATCATCAGTCG 1378

Db 241 ACCACACACATCATCAGTCG 261

|||||

RESULT 22

ABA64860

ID ABA64860 standard; DNA; 261 BP.

XX AC ABA64860;

XX

XX 01-FEB-2002 (first entry)

DT Human foetal liver single exon nucleic acid probe #13165.

DE Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

XX Homo sapiens.

OS

XX WO200157277-A2.

PN

XX 09-AUG-2001.

PD

XX 30-JAN-2001; 2001WO-US00669.

PF

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX

XX (MOLE-) MOLECULAR DYNAMICS INC.

PA Penn SG, Hanzel DK, Chen W, Rank DR;

PI WPI; 2001-483447/52.

XX

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human fetal liver -

PT Claim 4; SEQ ID NO 13165; 639pp + sequence listing; English.

PS

XX The invention relates to a single exon nucleic acid probe for

XX measuring human gene expression in a sample derived from human foetal

XX liver. The single exon nucleic acid probes may be used for predicting,

XX measuring and displaying gene expression in samples derived from human

XX fetal liver. The present sequence is a single exon nucleic acid

XX probe of the invention.

CC Note: The sequence data for this patent did not form part of the

CC printed specification, but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX

XX Sequence 261 BP; 73 A; 73 C; 45 G; 70 T; 0 other;

SQ

Query Match 10.1%; Score 261; DB 22; Length 261;

Best Local Similarity 100.0%; Pred. No. 2.2e-64;

Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1118 GTGGAACAAGTCAATCAGATATTTCAACTCTTCTACGTCCTCCATCAAGTCTGATTGG 1177

Db 1 GTGGAACAAGTCAATCAGATATTTCAACTCTTCTACGTCCTCCATCAAGTCTGATTGG 60

Qy 1178 AAGTGAGTGAAGTCAATGGAAGTAGATATCTCCAGCTGAACAATTTCTTCAGCCTTCTA 1237

Db 61 AAGTGAGTGAAGTCAATGGAAGTAGATATCTCCAGCTGAACAATTTCTTCAGCCTTCTA 120

Qy 1238 CATCCTCTACAAATGTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCATT 1297

Db 121 CATCCTCTACAAATGTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCATT 180

Qy 1298 CTACTCCTTCTATCTTCTCCAGACAGTGACAAAGGCGAGTCTGTGGCATCTGGAC 1357

|||||

01-FEB-2002 (first entry)

Human breast cell single exon nucleic acid probe #5674.

Human; microarray; single exon probe; gene expression; breast;

disease; cancer; ss.

Homo sapiens.

WO200157271-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US00662.

04-FEB-2000; 2000US-0180312.

26-MAY-2000; 2000US-0207456.

30-JUN-2000; 2000US-0608408.

03-AUG-2000; 2000US-0632366.

21-SEP-2000; 2000US-0234687.

27-SEP-2000; 2000US-0236359.

04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-496933/54.

New spatially-addressable set of single exon nucleic acid probes,

useful for measuring gene expression in sample derived from human

breast, comprises number of single exon nucleic acid probes -

Claim 4; SEQ ID NO 5674; 327pp + sequence listing; English.

The invention relates to a spatially-addressable set of single exon

nucleic acid probes for measuring gene expression in a sample derived

from human breast and BT 474 cells. The method involves contacting

the probes with a collection of detectably labelled nucleic acids

derived from mRNA of human breast, and then measuring the label

bound to each probe of the microarray. The probes are useful for

verifying the expression of regions of genomic DNA predicted to

encode proteins. They are useful for gene discovery, and for

determining predisposition and/or prognosing breast disease. Gene

expression analysis is useful for assessing the toxicity of chemical

agents on cells. The microarray of this invention presents a far greater

diversity of probes for measuring gene expression, with far less bias

than expressed sequence tag microarrays. The method is suitable for

rapid production of functional information from genomic sequence. The

present sequence is a single exon nucleic acid probe of the invention.

Note: The sequence data for this patent did not form part of the

printed specification, but was obtained in electronic format directly

from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 261 BP; 73 A; 73 C; 45 G; 70 T; 0 other;

Query Match 10.1%; Score 261; DB 22; Length 261;

Best Local Similarity 100.0%; Pred. No. 2.2e-64;

Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1118 GTGGAACAAGTCAATCAGATATTTCAACTCTTCTACGTCCTCCATCAAGTCTGATTGG 1177

Db 1 GTGGAACAAGTCAATCAGATATTTCAACTCTTCTACGTCCTCCATCAAGTCTGATTGG 60

Qy 1178 AAGTGAGTGAAGTCAATGGAAGTAGATATCTCCAGCTGAACAATTTCTTCAGCCTTCTA 1237

Db 61 AAGTGAGTGAAGTCAATGGAAGTAGATATCTCCAGCTGAACAATTTCTTCAGCCTTCTA 120

Qy 1238 CATCCTCTACAAATGTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCATT 1297

Db 121 CATCCTCTACAAATGTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCATT 180

Qy 1298 CTACTCCTTCTATCTTCTCCAGACAGTGACAAAGGCGAGTCTGTGGCATCTGGAC 1357



Db 121 CATCTCTACAAATGTCAGCTCAGGCTCATTCGACATCATCTCCACAGAAAGCCCTCATT 180  
 QY 1298 CTACTCCCTTTGCTATCTCTCCACAGAGTGAACAAAGCCAGCTGTGTGAGGCAATCTGGAC 1357  
 Db 181 CTACTCCCTTTGCTATCTCTCCACAGAGTGAACAAAGCCAGCTGTGTGAGGCAATCTGGAC 240  
 QY 1358 ACCACACATCATCATCTG 1378  
 Db 241 ACCACACATCATCATCTG 261

RESULT 23  
 ABA31975  
 ID ABA31975 standard; DNA; 261 BP.  
 XX AC ABA31975;  
 XX DT 23-JAN-2002 (first entry)  
 XX DE Probe #10441 for gene expression analysis in human heart cell sample.  
 XX KW Human; gene expression; heart; microarray; vascular system; probe;  
 KW cardiovascular disease; hypertension; cardiac arrhythmia;  
 XX KW congenital heart disease; ss.  
 XX OS Homo sapiens.  
 XX PN WO200157274-A2.  
 XX PD 09-AUG-2001.  
 XX PF 30-JAN-2001; 2001WO-US00666.  
 XX PR 04-FEB-2000; 2000US-0180312.  
 XX PR 26-MAY-2000; 2000US-0207456.  
 XX PR 30-JUN-2000; 2000US-0608408.  
 XX PR 03-AUG-2000; 2000US-0632366.  
 XX PR 21-SEP-2000; 2000US-0234687.  
 XX PR 27-SEP-2000; 2000US-0236359.  
 XX PR 04-OCT-2000; 2000GB-0024263.  
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX DR WPI; 2001-488899/53.  
 XX PT Single exon nucleic acid probes for analyzing gene expression in human hearts -  
 XX PS Claim 4; SEQ ID No 10441; 530pp; English.  
 XX CC The present invention relates to single exon nucleic acid probes for measuring human gene expression in a sample derived from human heart. The present sequence is one such probe. The probes may be used for predicting, measuring and displaying gene expression in samples derived from the human heart via microarrays. By measuring gene expression, the probes are useful for predicting, diagnosing, grading, staging, monitoring and prognosing diseases of the human heart and vascular system e.g. cardiovascular disease, hypertension, cardiac arrhythmias and congenital heart disease.  
 XX CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pt\_sequences.  
 XX CC Sequence 261 BP; 73 A; 73 C; 45 G; 70 T; 0 other;  
 XX CC Query Match 10.1%; Score 261; DB 22; Length 261;  
 XX CC Best Local Similarity 100.0%; Pred. No. 2.2e-64;  
 XX CC Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1118 GTGCAACAGTCAATCAGATATTCACTTCTCCTACGGTCCCATCAAGTCTGATTGG 1177  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 1 GTGGAACAGTCAATCAGATATTCACTTCTCCTACGGTCCCATCAAGTCTGATTGG 60  
 QY 1178 AAGTGAGTGAAGCTGCAATGGAGTAGATACCTCCAGCTGAACAAATTTCTTTCAGCCTTCTA 1237  
 Db 61 AAGTGAGTGAAGCTGCAATGGAGTAGATACCTCCAGCTGAACAAATTTCTTTCAGCCTTCTA 120  
 QY 1238 CATCTCTACAAATGTCAGCTCAGGCTCATTCGACATCATCTCCACAGAAAGCCCTCATT 1297  
 Db 121 CATCTCTACAAATGTCAGCTCAGGCTCATTCGACATCATCTCCACAGAAAGCCCTCATT 180  
 QY 1298 CTACTCCCTTTGCTATCTCTCCACAGAGTGAACAAAGCCAGCTGTGTGAGGCAATCTGGAC 1357  
 Db 181 CTACTCCCTTTGCTATCTCTCCACAGAGTGAACAAAGCCAGCTGTGTGAGGCAATCTGGAC 240  
 QY 1358 ACCACACATCATCATCTG 1378  
 Db 241 ACCACACATCATCATCTG 261

RESULT 24  
 AAK13288  
 ID AAK13288 standard; DNA; 261 BP.  
 XX AC AAK13288;  
 XX DT 05-NOV-2001 (first entry)  
 XX DE Human brain expressed single exon probe SEQ ID NO: 13279.  
 XX KW Human; brain expressed exon; gene expression analysis; probe;  
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
 XX KW epilepsy; cancer; ss.  
 XX OS Homo sapiens.  
 XX PN WO200157275-A2.  
 XX PD 09-AUG-2001.  
 XX PF 30-JAN-2001; 2001WO-US00667.  
 XX PR 04-FEB-2000; 2000US-0180312.  
 XX PR 26-MAY-2000; 2000US-0207456.  
 XX PR 30-JUN-2000; 2000US-0608408.  
 XX PR 03-AUG-2000; 2000US-0632366.  
 XX PR 21-SEP-2000; 2000US-0234687.  
 XX PR 27-SEP-2000; 2000US-0236359.  
 XX PR 04-OCT-2000; 2000GB-0024263.  
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX DR WPI; 2001-483446/52.  
 XX PT Single exon nucleic acid probes for analyzing gene expression in human brains -  
 XX PS Example 4; SEQ ID NO: 13279; 650pp + Sequence Listing; English.  
 XX CC The present invention provides a number of single exon nucleic acid probes which are derived from genomic sequences expressed in the human brain. They can be used to measure gene expression in brain cell samples, which may enable the diagnosis and improved treatment of nervous system diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia, epilepsy and cancers. The present sequence is one of the probes of the invention.  
 XX CC Sequence 261 BP; 73 A; 73 C; 45 G; 70 T; 0 other;  
 XX CC Query Match 10.1%; Score 261; DB 22; Length 261;  
 XX CC Best Local Similarity 100.0%; Pred. No. 2.2e-64;  
 XX CC Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches	261;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
y	1118	GTGGAACAAGTCAATCAGATATTTCAACTCTTCTACGGTCCCATCAAGTCTCGATTGG	1177						
b	1	GTGGAACAAGTCAATCAGATATTTCAACTCTTCTACGGTCCCATCAAGTCTCGATTGG	60						
y	1178	AAGTGAGTGAATCGAATGGAAGTAGATACTCCAGCTGAACAATTTCTTCAGCCTTCTA	1237						
b	61	AAGTGAGTGAATCGAATGGAAGTAGATACTCCAGCTGAACAATTTCTTCAGCCTTCTA	120						
y	1238	CATCCCTCAAAATGTCAGCTCAGGCCTCATTCGACATCATCTCCCACAAAAAGCCCTCAT	1297						
b	121	CATCCCTCAAAATGTCAGCTCAGGCCTCATTCGACATCATCTCCCACAAAAAGCCCTCAT	180						
y	1298	CTACTCCTTTGCTATCTTTCTCCAGACAGTGAACAAGGCAGTCTGTTGAGGCATCTGGAC	1357						
b	181	CTACTCCTTTGCTATCTTTCTCCAGACAGTGAACAAGGCAGTCTGTTGAGGCATCTGGAC	240						
y	1358	ACCACACATCATCAGTCTG	1378						
b	241	ACCACACATCATCAGTCTG	261						

RESULT 26  
AAI19833  
ID AAI19833 standard; DNA; 261 BP.

XX AC AAI19833;  
XX AC  
DT 12-OCT-2001 (first entry)  
DE Probe #9766 for gene expression analysis in human cervical cell sample.  
KW Probe; human; microarray; gene expression; cervical epithelial cell;  
KW cervical cancer; ss.  
XX Homo sapiens.  
OS OS  
PN WO200157278-A2.  
PD 09-AUG-2001.  
XX 30-JAN-2001; 2001WO-US006070.  
PF 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
PA Penn SG, Hanzel DK, Chen W, Rank DR;  
PI WPI; 2001-488901/53.  
XX Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human cervical epithelial cells -  
PS Claim 25; SEQ ID No 9766; 487pp; English.  
XX The present invention relates to human single exon nucleic acid probes  
CC (SENP). The present sequence is one such probe. The SENPs are derived  
CC from human HeLa cells. The SENPs can be used to produce a single exon  
CC microarray, which can be used for measuring human gene expression in a  
CC sample derived from human cervical epithelial cells. By measuring gene  
CC expression, the probes are therefore useful in grading and/or staging  
CC of diseases of the cervix, notably cervical cancer.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

QY 1118 GTGGACAGTCAATCAGATATTCAACTCTTCCTACGGTCCCATCAAGTCTGATTGG 1177  
Db 1 GTGGAAACAAGTCAATCAGATATTCAACTCTTCCTACGGTCCCATCAAGTCTGATTGG 60

QY 1178 AAGTGAGTGAACACTGCAATGGAAGTAGAATACTCCAGCTGAACAATTTCTTCAGCCCTTCTA 1237  
Db 61 AAGTGAGTGAACACTGCAATGGAAGTAGAATACTCCAGCTGAACAATTTCTTCAGCCCTTCTA 120

QY 1238 CATCCTCTACAAATGCAGCTCAGGCTCATTCGACATCATCTCCCACAGAAGCCCTCATTT 1297  
Db 121 CATCCTCTACAAATGCAGCTCAGGCTCATTCGACATCATCTCCCACAGAAGCCCTCATTT 180

QY 1298 CTACTCCTTTGTATCTTCTCCAGACAGTGAACAAGCGAGTCTGTTGAGGCATCTGGAC 1357  
Db 181 CTACTCCTTTGTATCTTCTCCAGACAGTGAACAAGCGAGTCTGTTGAGGCATCTGGAC 240

QY 1358 ACCACACATCATCATCAGTCTG 1378  
Db 241 ACCACACATCATCATCAGTCTG 261

RESULT 25  
AAK39023  
ID AAK39023 standard; DNA; 261 BP.  
AC AAK39023;  
DT 06-NOV-2001 (first entry)  
DE Human bone marrow expressed single exon probe SEQ ID NO: 13580.  
KW Human; bone marrow expressed exon; gene expression analysis; probe;  
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.  
XX Homo sapiens.  
XX WO200157276-A2.  
PN 09-AUG-2001.  
XX 30-JAN-2001; 2001WO-US00668.  
XX 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.  
PA Penn SG, Hanzel DK, Chen W, Rank DR;  
PI WPI; 2001-488900/53.  
DR Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human bone marrow -  
PS Example 4; SEQ ID NO: 13580; 659pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid  
XX probes which are derived from genomic sequences expressed in the human  
CC bone marrow. They can be used to measure gene expression in bone marrow  
CC samples, which may enable the improved diagnosis and treatment of cancers  
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of  
CC the probes of the invention.

XX Sequence 261 BP; 73 A; 73 C; 45 G; 70 T; 0 other;  
SQ Query Match 10.1%; Score 261; DB 22; Length 261;  
Best Local Similarity 100.0%; Pred. No. 2.2e-64;

XX SQ Sequence 261 BP; 73 A; 73 C; 45 G; 70 T; 0 other;  
 Query Match 10.1%; Score 261; DB 22; Length 261;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-64;  
 Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1118 GTGGAACAAGTCAATCAGATATTTCAACTCTTCTAGGGTCCCATCAAGTCCCTGATTGG 1177  
 Db 1 GTGGAACAAGTCAATCAGATATTTCAACTCTTCTAGGGTCCCATCAAGTCCCTGATTGG 60

QY 1178 AAGTGAGTGAAGTCAATGGAAGTAGATCTCCAGCTGAACAATTTCTTCAGCCTTCTA 1237  
 Db 61 AAGTGAGTGAAGTCAATGGAAGTAGATCTCCAGCTGAACAATTTCTTCAGCCTTCTA 120

QY 1238 CATCTCTACAAATGTCAGCTCAGGCTCAATTCGACATCATCTCCACAGAAAGCCCTCATT 1297  
 Db 121 CATCTCTACAAATGTCAGCTCAGGCTCAATTCGACATCATCTCCACAGAAAGCCCTCATT 180

QY 1298 CTACTCCTTTGCTATCTTCTCCAGACAGTGAACAAGGAGTCTTTGAGGCATCTGGAC 1357  
 Db 181 CTACTCCTTTGCTATCTTCTCCAGACAGTGAACAAGGAGTCTTTGAGGCATCTGGAC 240

QY 1358 ACCACACATCATCAGTCTG 1378  
 Db 241 ACCACACATCATCAGTCTG 261

RESULT 27  
 AAI45028  
 ID AAI45028 standard; DNA; 261 BP.  
 AC AAI45028;  
 DT 17-OCT-2001 (first entry)  
 DE Probe #13714 used to measure gene expression in human placenta sample.  
 KW Probe; microarray; human; placenta; antenatal diagnosis;  
 KW genetic disorder; ss.  
 XX Homo sapiens.  
 OS  
 PN WO200157272-A2.  
 XX 09-AUG-2001.  
 XX 30-JAN-2001; 2001WO-US00663.  
 XX 04-FEB-2000; 2000US-0180312.  
 XX 26-MAY-2000; 2000US-0207456.  
 XX 30-JUN-2000; 2000US-0608408.  
 XX 03-AUG-2000; 2000US-0632366.  
 XX 21-SEP-2000; 2000US-0234687.  
 XX 27-SEP-2000; 2000US-0236359.  
 XX 04-OCT-2000; 2000GB-0024263.  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2001-488897/53.  
 XX Human genome-derived single exon nucleic acid probes useful for  
 XX analyzing gene expression in human placenta -  
 PS Claim 25; SEQ ID No 13714; 654pp; English.

CC The present invention relates to single exon nucleic acid probes (SENP).  
 CC The present sequence is one such probe. The probes are useful for  
 CC producing a microarray for predicting, measuring and displaying gene  
 CC expression in samples derived from human placenta. The probes are useful  
 CC for antenatal diagnosis of human genetic disorders.

XX SQ Sequence 261 BP; 73 A; 73 C; 45 G; 70 T; 0 other;  
 Query Match 10.1%; Score 261; DB 22; Length 261;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-64;  
 Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1118 GTGGAACAAGTCAATCAGATATTTCAACTCTTCTAGGGTCCCATCAAGTCCCTGATTGG 1177  
 Db 1 GTGGAACAAGTCAATCAGATATTTCAACTCTTCTAGGGTCCCATCAAGTCCCTGATTGG 60

QY 1178 AAGTGAGTGAAGTCAATGGAAGTAGATCTCCAGCTGAACAATTTCTTCAGCCTTCTA 1237  
 Db 61 AAGTGAGTGAAGTCAATGGAAGTAGATCTCCAGCTGAACAATTTCTTCAGCCTTCTA 120

QY 1238 CATCTCTACAAATGTCAGCTCAGGCTCAATTCGACATCATCTCCACAGAAAGCCCTCATT 1297  
 Db 121 CATCTCTACAAATGTCAGCTCAGGCTCAATTCGACATCATCTCCACAGAAAGCCCTCATT 180

QY 1298 CTACTCCTTTGCTATCTTCTCCAGACAGTGAACAAGGAGTCTTTGAGGCATCTGGAC 1357  
 Db 181 CTACTCCTTTGCTATCTTCTCCAGACAGTGAACAAGGAGTCTTTGAGGCATCTGGAC 240

QY 1358 ACCACACATCATCAGTCTG 1378  
 Db 241 ACCACACATCATCAGTCTG 261

RESULT 28  
 AAI05549  
 ID AAI05549 standard; DNA; 261 BP.  
 AC AAI05549;  
 DT 09-OCT-2001 (first entry)  
 DE Probe #5540 used to measure gene expression in human breast sample.  
 KW Probe; human; breast disease; breast cancer; development disorder; ss;  
 KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.  
 XX Homo sapiens.  
 OS  
 PN WO200157270-A2.  
 XX 09-AUG-2001.  
 XX 29-JAN-2001; 2001WO-US00661.  
 XX 04-FEB-2000; 2000US-0180312.  
 XX 26-MAY-2000; 2000US-0207456.  
 XX 30-JUN-2000; 2000US-0608408.  
 XX 03-AUG-2000; 2000US-0632366.  
 XX 21-SEP-2000; 2000US-0234687.  
 XX 27-SEP-2000; 2000US-0236359.  
 XX 04-OCT-2000; 2000GB-0024263.  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2001-476286/51.  
 XX Novel, single exon nucleic acid probe used to measuring gene expression  
 XX in a human breast -  
 PS Claim 25; SEQ ID No 5540; 322pp; English.

CC The present invention relates to novel single exon nucleic acid probes.  
 CC The present sequence is one such probe. The probes are useful for  
 CC measuring human gene expression in a human breast sample, where the probe  
 CC hybridises at high stringency to a nucleic acid expressed in the human  
 CC breast. The probes are useful for predicting, diagnosing, grading,

staging, monitoring and prognosing diseases of the human breast,  
particularly those diseases with polygenic aetiology. The diseases  
include: breast cancer, disorders of development, inflammatory diseases  
of the breast, fibrocystic changes, proliferative breast disease and  
non-carcinoma tumours.

Note: The sequence data for this patent did not form part of the printed  
specification, but was obtained in electronic format directly from WIPO  
at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 261 BP; 73 A; 73 C; 45 G; 70 T; 0 other;  
SQ

Query Match 10.1%; Score 261; DB 22; Length 261;  
Best Local Similarity 100.0%; Pred. No. 2.2e-64;  
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1118 GTGGAAACAGTCAATCAGATATTTCACCTCTTCCTACGGTCCCATCAAGTCCTGATTGG 1177  
DB 1 GTGGAAACAGTCAATCAGATATTTCACCTCTTCCTACGGTCCCATCAAGTCTGATTGG 60

QY 1178 AAGTGAGTGAAGTGCATGAGTAGACTGCCAGCTGAACAATTTCTTCAGCCTTCTA 1237  
DB 61 AAGTGAGTGAAGTGCATGAGTAGACTGCCAGCTGAACAATTTCTTCAGCCTTCTA 120

QY 1238 CATCCTCTACAATGTCAAGTCAGCTCAGGCTCATTCACATCATCTCCACAGAAGCCCTCAT 1297  
DB 121 CATCCTCTACAATGTCAAGTCAGCTCAGGCTCATTCACATCATCTCCACAGAAGCCCTCAT 180

QY 1298 CTACTCCTTTGCTATCTTCTCCACAGAGTGAACAAGGCGAGTCTGTTGAGGCATCTGGAC 1357  
DB 181 CTACTCCTTTGCTATCTTCTCCACAGAGTGAACAAGGCGAGTCTGTTGAGGCATCTGGAC 240

QY 1358 ACCACACATCATCAGTCTG 1378  
DB 241 ACCACACATCATCAGTCTG 261

RESULT 29  
ABS13107 standard; DNA; 261 BP.  
XX AC ABS13107;  
XX DT  
XX 19-AUG-2002 (first entry)  
XX Human genome-derived single exon probe ORF from lung SEQ ID No 13098.  
XX Human; ds; single exon probe; asthma; lung cancer; COPD; IID;  
KW chronic obstructive pulmonary disease; interstitial lung disease;  
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;  
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
KW primary ciliary dyskinesia; pulmonary hypertension;  
KW hyaline membrane disease; open reading frame; ORF.  
XX OS  
XX Homo sapiens.  
PN WO200186003-A2.  
XX PD  
XX 15-NOV-2001.  
XX PF  
XX 30-JAN-2001; 2001WO-US00665.  
XX QY  
XX 04-FEB-2000; 2000US-180312P.  
PR 26-MAY-2000; 2000US-207456P.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-234687P.  
PR 27-SEP-2000; 2000US-236359P.  
PR 04-OCT-2000; 2000GB-0024263.  
XX PA  
XX (MOLE-) MOLECULAR DYNAMICS INC.

QY 1358 ACCACACATCATCAGTCTG 1378  
 DB 241 ACCACACATCATCAGTCTG 261

RESULT 30  
 ID ABA43873 standard; DNA; 481 BP.  
 AC ABA43873;  
 DT 01-FEB-2002 (first entry)  
 DE Human breast cell single exon nucleic acid probe #2568.  
 KW Human; microarray; single exon probe; gene expression; breast;  
 KW disease; cancer; ss.  
 XX Homo sapiens.  
 OS  
 PN WO200157271-A2.  
 PD 09-AUG-2001.  
 PF 30-JAN-2001; 2001WO-US006562.  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 PI WPI; 2001-496933/54.  
 PT New spatially-addressable set of single exon nucleic acid probes,  
 PT useful for measuring gene expression in sample derived from human  
 PT breast, comprises number of single exon nucleic acid probes -  
 PS Claim 1; SEQ ID NO 2568; 327pp + sequence listing; English.  
 XX The invention relates to a spatially-addressable set of single exon  
 XX nucleic acid probes for measuring gene expression in a sample derived  
 XX from human breast and BT 474 cells. The method involves contacting  
 XX the probes with a collection of detectably labelled nucleic acids  
 XX derived from mRNA of human breast, and then measuring the label  
 XX bound to each probe of the microarray. The probes are useful for  
 XX verifying the expression of regions of genomic DNA predicted to  
 XX encode proteins. They are useful for gene discovery, and for  
 XX determining predisposition and/or prognosing breast disease. Gene  
 XX expression analysis is useful for assessing the toxicity of chemical  
 XX agents on cells. The microarray of this invention presents a far greater  
 XX diversity of probes for measuring gene expression, with far less bias  
 XX than expressed sequence tag microarrays. The method is suitable for  
 XX rapid production of functional information from genomic sequence. The  
 XX present sequence is a single exon nucleic acid probe of the invention.  
 XX Note: The sequence data for this patent did not form part of the  
 XX printed specification, but was obtained in electronic format directly  
 XX from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 SQ Sequence 481 BP; 152 A; 106 C; 89 G; 134 T; 0 other;

Query Match 9.9%; Score 255.8; DB 22; Length 481;  
 Best Local Similarity 99.2%; Pred. No. 9.4e-63;  
 Matches 257; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1696 ATAACAATCAGATAGTTCAGCCAGCCATTGGATTCCAACTCAGGAGAAAGAAAT 1755  
 ID ABA43873 standard; DNA; 481 BP.  
 AC ABA43873;  
 DT 01-FEB-2002 (first entry)  
 DE Human breast cell single exon nucleic acid probe #2568.  
 KW Human; microarray; single exon probe; gene expression; breast;  
 KW disease; cancer; ss.  
 XX Homo sapiens.  
 OS  
 PN WO200157271-A2.  
 PD 09-AUG-2001.  
 PF 30-JAN-2001; 2001WO-US006562.  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 PI WPI; 2001-488895/53.  
 PT Single exon nucleic acid probes for analyzing gene expression in human  
 PT hearts -  
 PS Claim 1; SEQ ID NO 2560; 530pp; English.  
 XX The present invention relates to single exon nucleic acid probes for  
 XX measuring human gene expression in a sample derived from human heart. The  
 XX present sequence is one such probe. The probes may be used for  
 XX predicting, measuring and displaying gene expression in samples derived  
 XX from the human heart via microarrays. By measuring gene expression, the  
 XX probes are useful for predicting, diagnosing, grading, staging,  
 XX monitoring and prognosing diseases of the human heart and vascular system  
 XX e.g. cardiovascular disease, hypertension, cardiac arrhythmias and  
 XX congenital heart disease.  
 XX Note: The sequence data for this patent did not form part of the printed  
 XX specification, but was obtained in electronic format directly from WIPO  
 XX at ftp.wipo.int/pub/published\_pct\_sequences.

DB 223 ACAGCACATCAGATAAGTTCACAGCCAGCCATTGGATTCCAACTCAGGAGAAAGAAAT 282  
 QY 1756 GACCTCAATCTTGATCGCTCTTTGGGGTTCAGAGAAATCTGCTTCATCTCAAAAAGCC 1815  
 DB 283 GACCTCAATCTTGATCGCTCTTTGGGGTTCAGAGAAATCTGCTTCATCTCAAAAAGCC 342  
 QY 1816 AAGGAACCCAGAACTTCAGATCAGACTAGCAGTCTGAGAGTGTACCAATGAAATAACACC 1875  
 DB 343 AAGGAACCCAGAACTTCAGATCAGACTAGCAGTCTGAGAGTGTACCAATGAAATAACACC 402  
 QY 1876 AATCTGAGCTCTGAGTCTTCCAAACAGAGCCACTGGGCTTCAGCTCATGAGAAACATCC 1935  
 DB 403 AATCTGAGCTCTGAGTCTTCCAAACAGAGCCACTGGGCTTCAGCTCATGAGAAACATCC 462  
 QY 1936 ACCAGGACTCTGCTCTTC 1954  
 DB 463 ACCAGGACTCTGCTCTTC 481

RESULT 31  
 ID ABA24094 standard; DNA; 481 BP.  
 AC ABA24094;  
 DT 23-JAN-2002 (first entry)  
 DE Probe #2560 for gene expression analysis in human heart cell sample.  
 KW Human; gene expression; heart; microarray; vascular system; probe;  
 KW cardiovascular disease; hypertension; cardiac arrhythmia;  
 KW congenital heart disease; ss.  
 XX Homo sapiens.  
 OS  
 PN WO200157274-A2.  
 PD 09-AUG-2001.  
 PF 30-JAN-2001; 2001WO-US006566.  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 PI WPI; 2001-488895/53.  
 PT Single exon nucleic acid probes for analyzing gene expression in human  
 PT hearts -  
 PS Claim 1; SEQ ID NO 2560; 530pp; English.  
 XX The present invention relates to single exon nucleic acid probes for  
 XX measuring human gene expression in a sample derived from human heart. The  
 XX present sequence is one such probe. The probes may be used for  
 XX predicting, measuring and displaying gene expression in samples derived  
 XX from the human heart via microarrays. By measuring gene expression, the  
 XX probes are useful for predicting, diagnosing, grading, staging,  
 XX monitoring and prognosing diseases of the human heart and vascular system  
 XX e.g. cardiovascular disease, hypertension, cardiac arrhythmias and  
 XX congenital heart disease.  
 XX Note: The sequence data for this patent did not form part of the printed  
 XX specification, but was obtained in electronic format directly from WIPO  
 XX at ftp.wipo.int/pub/published\_pct\_sequences.

us-09-781-693a-1.l.rng

Wed Mar 26 09:38:49 2003

CC epilepsy and cancers. The present sequence is one of the probes of the  
 CC invention.  
 SQ Sequence 481 BP; 152 A; 106 C; 89 G; 134 T; 0 other;

Query Match 9.9%; Score 255.8; DB 22; Length 481;  
 Best Local Similarity 99.2%; Pred. No. 9.4e-63;  
 Matches 257; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1696 ATACACAAATCAGATAAGTTCCACAGCCCAAGCCATTTGGATTCCAACTCAGGAGAGAAAT 1755  
 DB 223 ACAGCACAATCAGATAAGTTCCACAGCCCAAGCCATTTGGATTCCAACTCAGGAGAGAAAT 282  
 QY 1756 GACCTCAATCTTGATCGCTCTTGTGGGTTCCAGAGAAATCTGCTTCATCTGAAAAGCC 1815  
 DB 283 GACCTCAATCTTGATCGCTCTTGTGGGTTCCAGAGAAATCTGCTTCATCTGAAAAGCC 342  
 QY 1816 AAGGAACCAAGAACTTCAGATCAGACTAGCTAGAGTGTACCAATGAAAATTAACACC 1875  
 DB 343 AAGGAACCAAGAACTTCAGATCAGACTAGCTAGAGTGTACCAATGAAAATTAACACC 402  
 QY 1876 AATCTTGAGCTCAGTTCCAAACAGAGCCACTGGGCTTCAGCTCATGAAAGAAATCC 1935  
 DB 403 AATCTTGAGCTCAGTTCCAAACAGAGCCACTGGGCTTCAGCTCATGAAAGAAATCC 462  
 QY 1936 ACCAGGACTCTGCTCTTC 1954  
 DB 463 ACCAGGACTCTGCTCTTC 481

RESULT 32  
 AAK02620  
 ID AAK02620 standard; DNA; 481 BP.  
 XX  
 AC AAK02620;  
 DT  
 DT  
 DT  
 DE  
 DE  
 DE  
 KW Human brain expressed single exon probe SEQ ID NO: 2611.  
 KW Human; brain expressed exon; gene expression analysis; probe;  
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
 KW epilepsy; cancer; ss.  
 XX Homo sapiens.  
 OS  
 OS  
 PN WO200157275-A2.  
 PN  
 PD 09-AUG-2001.  
 PD  
 PF 30-JAN-2001; 2001WO-US00667.  
 PF  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0623366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 XX (MOLE-) MOLECULAR DYNAMICS INC.

PA Penn SG, Hanzel DK, Chen W, Rank DR;  
 PI WPI; 2001-483446/52.  
 XX Single exon nucleic acid probes for analyzing gene expression in human  
 XX brains -  
 XX Example 4; SEQ ID NO: 2611; 650pp + Sequence Listing; English.  
 XX The present invention provides a number of single exon nucleic acid  
 XX probes which are derived from genomic sequences expressed in the human  
 XX brain. They can be used to measure gene expression in brain cell samples,  
 XX which may enable the diagnosis and improved treatment of nervous system  
 XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
 CC

QY 1696 ATACACAAATCAGATAAGTTCCACAGCCCAAGCCATTTGGATTCCAACTCAGGAGAGAAAT 1755  
 DB 223 ACAGCACAATCAGATAAGTTCCACAGCCCAAGCCATTTGGATTCCAACTCAGGAGAGAAAT 282  
 QY 1756 GACCTCAATCTTGATCGCTCTTGTGGGTTCCAGAGAAATCTGCTTCATCTGAAAAGCC 1815  
 DB 283 GACCTCAATCTTGATCGCTCTTGTGGGTTCCAGAGAAATCTGCTTCATCTGAAAAGCC 342  
 QY 1816 AAGGAACCAAGAACTTCAGATCAGACTAGCTAGAGTGTACCAATGAAAATTAACACC 1875  
 DB 343 AAGGAACCAAGAACTTCAGATCAGACTAGCTAGAGTGTACCAATGAAAATTAACACC 402  
 QY 1876 AATCTTGAGCTCAGTTCCAAACAGAGCCACTGGGCTTCAGCTCATGAAAGAAATCC 1935  
 DB 403 AATCTTGAGCTCAGTTCCAAACAGAGCCACTGGGCTTCAGCTCATGAAAGAAATCC 462  
 QY 1936 ACCAGGACTCTGCTCTTC 1954  
 DB 463 ACCAGGACTCTGCTCTTC 481

RESULT 33  
 AAK28054  
 ID AAK28054 standard; DNA; 481 BP.  
 XX  
 AC AAK28054;  
 DT  
 DT  
 DT  
 DE  
 DE  
 DE  
 KW Human bone marrow expressed single exon probe SEQ ID NO: 2611.  
 KW Human; bone marrow expressed exon; gene expression analysis; probe;  
 KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.  
 XX Homo sapiens.  
 OS  
 OS  
 PN WO200157276-A2.  
 PN  
 PD 09-AUG-2001.  
 PD  
 PF 30-JAN-2001; 2001WO-US00668.  
 PF  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0623366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 XX (MOLE-) MOLECULAR DYNAMICS INC.

PA Penn SG, Hanzel DK, Chen W, Rank DR;  
 PI WPI; 2001-488900/53.  
 XX Human genome-derived single exon nucleic acid probes useful for  
 XX analyzing gene expression in human bone marrow -  
 XX Example 4; SEQ ID NO: 2611; 658pp + Sequence Listing; English.  
 XX The present invention provides a number of single exon nucleic acid  
 XX probes which are derived from genomic sequences expressed in the human  
 XX bone marrow. They can be used to measure gene expression in bone marrow  
 CC



us-09-781-693a-1.rng

Wed Mar 26 09:38:49 2003

XX PS Claim 25; SEQ ID No 2528; 322pp; English.

XX CC The present invention relates to novel single exon nucleic acid probes.

XX CC The present sequence is one such probe. The probes are useful for

XX CC measuring human gene expression in a human breast sample, where the probe

XX CC hybridises at high stringency to a nucleic acid expressed in the human

XX CC breast. The probes are useful for predicting, diagnosing, grading,

XX CC staging, monitoring and prognosing diseases of the human breast,

XX CC particularly those diseases with polygenic aetiology. The diseases

XX CC include: breast cancer, disorders of development, inflammatory disease and

XX CC of the breast, fibrocystic changes, proliferative breast disease and

XX CC non-carcinoma tumours.

XX CC Note: The sequence data for this patent did not form part of the printed

XX CC specification, but was obtained in electronic format directly from WIPO

XX CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 481 BP; 152 A; 106 C; 89 G; 134 T; 0 other;

Query Match 9.9%; Score 255.8; DB 22; Length 481;

Best Local Similarity 99.2%; Pred. No. 9.4e-63;

Matches 257; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1696 ATACCAATCAGATAGTTCACAGCCAGCCATTGGATTCCAACTCAGGAGAAAGAAAT 1755

Db 223 ACAGCAATCAGATAGTTCACAGCCAGCCATTGGATTCCAACTCAGGAGAAAGAAAT 282

Qy 1756 GACCTCAATCTTGATCGCTCTTGTGGGGTCCAGAGAAATCTGCTTCATCTGAAAGGCC 1815

Db 283 GACCTCAATCTTGATCGCTCTTGTGGGGTCCAGAGAAATCTGCTTCATCTGAAAGGCC 342

Qy 1816 AGGACACAGAAATCTCAGATCAGACTAGCAGTGTGCTTCCATGATGAAATACACC 1875

Qy 343 AGGACACAGAAATCTCAGATCAGACTAGCAGTGTGCTTCCATGATGAAATACACC 402

Qy 1876 AATCTTGAGCTCAGTTCACAGAGCCAGCCAGTGTGCTTCCATGATGAAATACACC 1935

Db 403 AATCTTGAGCTCAGTTCACAGAGCCAGCCAGTGTGCTTCCATGATGAAATACACC 462

Qy 1936 ACCAGGACTCTGCTCTTC 1954

Db 463 ACCAGGACTCTGCTCTTC 481

RESULT 36

ABS02539

ID ABS02539 standard; DNA; 481 BP.

XX AC ABS02539;

XX DT 19-AUG-2002 (first entry)

XX DE Human genome-derived single exon probe from lung SEQ ID No 2530.

XX KW Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;

XX KW chronic obstructive pulmonary disease; interstitial lung disease;

XX KW familial idiopathic pulmonary fibrosis; neurofibromatosis;

XX KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;

XX KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;

XX KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;

XX KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;

XX KW primary ciliary dyskinesia; pulmonary hypertension;

XX KW hyaline membrane disease.

XX OS Homo sapiens.

XX PN WO200186003-A2.

XX PD 15-NOV-2001.

XX PF 30-JAN-2001; 2001WO-US000665.

XX PS 04-FEB-2000; 2000US-180312P.

26-MAY-2000; 2000US-207456P.

30-JUN-2000; 2000US-0608408.

03-AUG-2000; 2000US-0632366.

21-SEP-2000; 2000US-234687P.

27-SEP-2000; 2000US-236359P.

04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2002-114183/15.

Spatially-addressable set of single exon nucleic acid probes, used to

measure gene expression in human lung samples -

Claim 1: SEQ ID No 2530; 634pp; English.

The invention relates to a spatially-addressable set of single exon

nucleic acid probes for measuring gene expression in a sample derived

from human lung comprising single exon nucleic acid probes having one of

12614 nucleic acid sequences mentioned in the specification, or their

complements or the 12387 open reading frames derived from the 12614

probes. Also included are a microarray comprising the novel set of

probes; the novel set of probes which hybridise at high stringency to a

nucleic acid expressed in the human lung; measuring gene expression in a

sample derived from human lung, comprising (a) contacting the array with

a collection of detectably labeled nucleic acids derived from human lung

mRNA, and (b) measuring the label detectably bound to each probe of

the array; identifying exons in a eukaryotic genome, comprising

(a) algorithmically predicting at least one exon from genomic sequences

of the eukaryote; and (b) detecting specific hybridisation of detectably

labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,

in the above mentioned microarray; assigning exons to a single gene,

comprising (a) identifying exons from genomic sequence by the method

above and (b) measuring the expression of each of the exons in several

tissues and/or cell types using hybridisation to a single exon

microarrays having a probe with the exon, where a common pattern of

expression of the exons in the tissues and/or cell types indicates that

the exons should be assigned to a single gene; a peptide comprising one

of 12011 sequences, mentioned in the specification, or encoded by the

expression analysis, and for identifying exons in a gene, particularly

using human lung derived mRNA and for the study of lung diseases

such as asthma, lung cancer, chronic obstructive pulmonary disease

(COPD), interstitial lung disease (ILD), familial idiopathic pulmonary

fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,

Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary

haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,

pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic

pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension

and hyaline membrane disease. The present sequence is a single exon

probe of the invention.

Note: The sequence data for this patent did not form part

of the printed specification, but was obtained in electronic

format directly from WIPO at

ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 481 BP; 152 A; 106 C; 89 G; 134 T; 0 other;

Query Match 9.9%; Score 255.8; DB 24; Length 481;

Best Local Similarity 99.2%; Pred. No. 9.4e-63;

Matches 257; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1696 ATACCAATCAGATAGTTCACAGCCAGCCATTGGATTCCAACTCAGGAGAAAGAAAT 1755

Db 223 ACAGCAATCAGATAGTTCACAGCCAGCCATTGGATTCCAACTCAGGAGAAAGAAAT 282

Qy 1756 GACCTCAATCTTGATCGCTCTTGTGGGGTCCAGAGAAATCTGCTTCATCTGAAAGGCC 1815

Db 283 GACCTCAATCTTGATCGCTCTTGTGGGGTCCAGAGAAATCTGCTTCATCTGAAAGGCC 342



QY 1816 AAGGAACCAAGAACTTCTAGATCAGACTAGACTGAGAGTCTACCAATGAAATAACAC 1875  
 Db 343 AAGGAACCAAGAACTTCTAGATCAGACTAGACTGAGAGTCTACCAATGAAATAACAC 402  
 QY 1876 AATCCTGAGCCTCAGTTCCTCAACAGAGCCACTGGCCTTCAGCTCATGGAAGAAATCC 1935  
 Db 403 AATCCTGAGCCTCAGTTCCTCAACAGAGCCACTGGCCTTCAGCTCATGGAAGAAATCC 462  
 QY 1936 ACCAGGACTCTGCTCTTC 1954  
 Db 463 ACCAGGACTCTGCTCTTC 481

RESULT 37  
 ABV60730  
 ID ABV60730 standard; cDNA; 423 BP.  
 AC ABV60730;  
 XX  
 DT 13-SEP-2002 (first entry)  
 XX  
 DE Human prostate expression marker cDNA 60721.  
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
 KW Pharmacogenomic marker; gene; ss.  
 OS Homo sapiens.  
 XX  
 PN WO200160860-A2.  
 XX  
 PD 23-AUG-2001.  
 XX  
 PF 20-FEB-2001; 2001WO-US05171.  
 XX  
 PR 17-FEB-2000; 2000US-183319P.  
 PR 16-MAR-2000; 2000US-189862P.  
 PR 25-MAY-2000; 2000US-207454P.  
 PR 03-JUN-2000; 2000US-211314P.  
 PR 18-JUL-2000; 2000US-219007P.  
 PR 13-DEC-2000; 2000US-255281P.  
 XX  
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
 PI Schlegel R, Endege WO, Monahan JE;  
 DR WPI; 2001-662795/76.  
 XX  
 PT Novel isolated nucleic acid molecule associated with cancerous state of  
 PT prostate cells and correlating with presence of prostate cancer, useful  
 PT for detecting presence of prostate cancer, stage of prostate cancer  
 XX  
 PS Claim 1; Page 11554-11555; 11750pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid molecule (I) comprising  
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
 CC specification or its complement. (I) is useful for:  
 CC (a) assessing whether a patient is afflicted with prostate cancer;  
 CC (b) monitoring the progression of prostate cancer in a patient;  
 CC (c) assessing the efficacy of a test compound to inhibit prostate  
 CC cancer in a patient;  
 CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
 CC in a patient;  
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
 CC (f) assessing the prostate cell carcinogenic potential of a compound;  
 CC (g) determining whether prostate cancer has metastasized in a patient;  
 CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
 CC patient;  
 CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.  
 XX  
 SQ Sequence 423 BP; 152 A; 87 C; 93 G; 91 T; 0 other;

Query Match 9.68; Score 248.4; DB 23; Length 423;  
 Best Local Similarity 97.7%; Pred. No. 1.2e-60;

Matches 252; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
 QY 1366 CATCATCAGTCTGATACAAATATGAAAGCTGAGCCCAACAGGACAGGTGAACCA 1425  
 Db 166 CATATCCGAGCTGATACAAATATGAAAGCTGAGCCCAACAGGACAGGTGAACCA 225  
 QY 1426 GTTTTAAGTTTGCACTACAGCAGAGGAGCAACTACAAGCAGACATATAAACTGAACCTT 1485  
 Db 226 GTTTTAAGTTTGCACTACAGCAGAGGAGCAACTACAAGCAGACATATAAACTGAACCTT 285  
 QY 1486 ACAGATGAATGGAGAGTATAGCATCAAGTTCTAGAGGAATTGGAGCCATTGCAATCT 1545  
 Db 286 ACAGATGAATGGAGAGTATAGCATCAAGTTCTAGAGGAATTGGAGCCATTGCAATCT 345  
 QY 1546 GAGGTCAGGAGGAATCTTTTCGTCACAGAGCTCAGTCAACCCAGAGAGGAGACAGT 1605  
 Db 346 GAGGTCAGGAGGAATCTTTTCGTCACAGAGCTCAGTCAACCCAGAGAGGAGACAGT 405  
 QY 1606 GAAACAAAAGCTCTCGAA 1623  
 Db 406 GAAACAAAAGCTCTCGAA 423

RESULT 38  
 ABA46980  
 ID ABA46980 standard; DNA; 186 BP.  
 AC ABA46980;  
 XX  
 DT 01-FEB-2002 (first entry)  
 XX  
 DE Human breast cell single exon nucleic acid probe #5675.  
 KW Human; microarray; single exon probe; gene expression; breast;  
 KW disease; cancer; ss.  
 OS Homo sapiens.  
 XX  
 PN WO200157271-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US00662.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 DR WPI; 2001-496933/54.  
 XX  
 PT New spatially-addressable set of single exon nucleic acid probes,  
 PT useful for measuring gene expression in sample derived from human  
 PT breast, comprises number of single exon nucleic acid probes  
 PS Claim 4; SEQ ID NO 5675; 327pp + sequence listing; English.  
 XX  
 CC The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human breast and Bt 474 cells. The method involves contacting  
 CC the probes with a collection of detectably labelled nucleic acids  
 CC derived from mRNA of human breast, and then measuring the label  
 CC bound to each probe of the microarray. The probes are useful for  
 CC verifying the expression of regions of genomic DNA predicted to  
 CC encode proteins. They are useful for gene discovery, and for  
 CC determining predisposition and/or prognosing breast disease. Gene

Wed Mar 26 09:38:49 2003

CC measuring human gene expression in a sample derived from human foetal  
CC liver. The single exon nucleic acid probes may be used for predicting,  
CC measuring and displaying gene expression in samples derived from human  
CC fetal liver. The present sequence is a single exon nucleic acid  
CC probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 186 BP; 63 A; 34 C; 36 G; 53 T; 0 other;

Query Match 7.2%; Score 186; DB 22; Length 186;  
Best Local Similarity 100.0%; Pred. No. 5.8e-43;  
Matches 186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 253 GTTTTGACAAACAAATTCGTTTCAGGGCCAGCAACATATTTAGTGCAAGTCTTACT 312  
DB 1 GTTTTGACAAACAAATTCGTTTCAGGGCCAGCAACATATTTAGTGCAAGTCTTACT 60  
QY 313 TGTACAAATGATAAACAAGATTTGATCTCTGGAGATGGAGTAATATTTATACCAAC 372  
DB 61 TGTACAAATGATAAACAAGATTTGATCTCTGGAGATGGAGTAATATTTATACCAAC 120  
QY 373 GTTGAGCAAGATGCAGAAACCAACAGACAAATGCCAATTTACGTGTCATTATGGAATCT 432  
DB 121 GTTGAGCAAGATGCAGAAACCAACAGACAAATGCCAATTTACGTGTCATTATGGAATCT 180  
QY 433 TATGAG 438  
DB 181 TATGAG 186

RESULT 40  
ABA64861  
ID ABA64861 standard; DNA; 186 BP.

XX ABA64861;  
XX 01-FEB-2002 (first entry)  
XX Human foetal liver single exon nucleic acid probe #13166.  
XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.  
XX Homo sapiens.  
XX WO200157277-A2.  
XX 09-AUG-2001.  
XX 30-JAN-2001; 2001WO-US00669.  
XX 04-FEB-2000; 2000US-0180312.  
XX 26-MAY-2000; 2000US-0207456.  
XX 30-JUN-2000; 2000US-0608408.  
XX 03-AUG-2000; 2000US-0632366.  
XX 21-SEP-2000; 2000US-0234687.  
XX 27-SEP-2000; 2000US-0236359.  
XX 04-OCT-2000; 2000GB-0024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-483447/52.  
XX Human genome-derived single exon nucleic acid probes useful for  
XX analyzing gene expression in human fetal liver -  
XX Claim 4; SEQ ID NO 13166; 639pp + sequence listing; English.  
XX The invention relates to a single exon nucleic acid probe for

CC expression analysis is useful for assessing the toxicity of chemical  
CC agents on cells. The microarray of this invention presents a far greater  
CC diversity of probes for measuring gene expression, with far less bias  
CC than expressed sequence tag microarrays. The method is suitable for  
CC rapid production of functional information from genomic sequences. The  
CC present sequence is a single exon nucleic acid probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 186 BP; 63 A; 34 C; 36 G; 53 T; 0 other;

Query Match 7.2%; Score 186; DB 22; Length 186;  
Best Local Similarity 100.0%; Pred. No. 5.8e-43;  
Matches 186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 253 GTTTTGACAAACAAATTCGTTTCAGGGCCAGCAACATATTTAGTGCAAGTCTTACT 312  
DB 1 GTTTTGACAAACAAATTCGTTTCAGGGCCAGCAACATATTTAGTGCAAGTCTTACT 60  
QY 313 TGTACAAATGATAAACAAGATTTGATCTCTGGAGATGGAGTAATATTTATACCAAC 372  
DB 61 TGTACAAATGATAAACAAGATTTGATCTCTGGAGATGGAGTAATATTTATACCAAC 120  
QY 373 GTTGAGCAAGATGCAGAAACCAACAGACAAATGCCAATTTACGTGTCATTATGGAATCT 432  
DB 121 GTTGAGCAAGATGCAGAAACCAACAGACAAATGCCAATTTACGTGTCATTATGGAATCT 180  
QY 433 TATGAG 438  
DB 181 TATGAG 186

RESULT 39  
ABA64861  
ID ABA64861 standard; DNA; 186 BP.

XX ABA64861;  
XX 01-FEB-2002 (first entry)  
XX Human foetal liver single exon nucleic acid probe #13166.  
XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.  
XX Homo sapiens.  
XX WO200157277-A2.  
XX 09-AUG-2001.  
XX 30-JAN-2001; 2001WO-US00669.  
XX 04-FEB-2000; 2000US-0180312.  
XX 26-MAY-2000; 2000US-0207456.  
XX 30-JUN-2000; 2000US-0608408.  
XX 03-AUG-2000; 2000US-0632366.  
XX 21-SEP-2000; 2000US-0234687.  
XX 27-SEP-2000; 2000US-0236359.  
XX 04-OCT-2000; 2000GB-0024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-483447/52.  
XX Human genome-derived single exon nucleic acid probes useful for  
XX analyzing gene expression in human fetal liver -  
XX Claim 4; SEQ ID NO 13166; 639pp + sequence listing; English.  
XX The invention relates to a single exon nucleic acid probe for

CC measuring human gene expression in a sample derived from human foetal  
CC liver. The single exon nucleic acid probes may be used for predicting,  
CC measuring and displaying gene expression in samples derived from human  
CC fetal liver. The present sequence is a single exon nucleic acid  
CC probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 186 BP; 63 A; 34 C; 36 G; 53 T; 0 other;

Query Match 7.2%; Score 186; DB 22; Length 186;  
Best Local Similarity 100.0%; Pred. No. 5.8e-43;  
Matches 186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 253 GTTTTGACAAACAAATTCGTTTCAGGGCCAGCAACATATTTAGTGCAAGTCTTACT 312  
DB 1 GTTTTGACAAACAAATTCGTTTCAGGGCCAGCAACATATTTAGTGCAAGTCTTACT 60  
QY 313 TGTACAAATGATAAACAAGATTTGATCTCTGGAGATGGAGTAATATTTATACCAAC 372  
DB 61 TGTACAAATGATAAACAAGATTTGATCTCTGGAGATGGAGTAATATTTATACCAAC 120  
QY 373 GTTGAGCAAGATGCAGAAACCAACAGACAAATGCCAATTTACGTGTCATTATGGAATCT 432  
DB 121 GTTGAGCAAGATGCAGAAACCAACAGACAAATGCCAATTTACGTGTCATTATGGAATCT 180  
QY 433 TATGAG 438  
DB 181 TATGAG 186

RESULT 40  
ABA64861  
ID ABA64861 standard; DNA; 186 BP.

XX ABA64861;  
XX 23-JAN-2002 (first entry)  
XX Probe #10442 for gene expression analysis in human heart cell sample.  
XX Human; gene expression; heart; microarray; vascular system; probe;  
XX cardiovascular disease; hypertension; cardiac arrhythmia;  
XX congenital heart disease; ss.  
XX Homo sapiens.  
XX WO200157274-A2.  
XX 09-AUG-2001.  
XX 30-JAN-2001; 2001WO-US00666.  
XX 04-FEB-2000; 2000US-0180312.  
XX 26-MAY-2000; 2000US-0207456.  
XX 30-JUN-2000; 2000US-0608408.  
XX 03-AUG-2000; 2000US-0632366.  
XX 21-SEP-2000; 2000US-0234687.  
XX 27-SEP-2000; 2000US-0236359.  
XX 04-OCT-2000; 2000GB-0024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488899/53.  
XX Single exon nucleic acid probes for analyzing gene expression in human  
XX hearts -  
XX Claim 4; SEQ ID No 10442; 530pp; English.

The present invention relates to single exon nucleic acid probes for measuring human gene expression in a sample derived from human heart. The present sequence is one such probe. The probes may be used for predicting, measuring and displaying gene expression in samples derived from the human heart via microarrays. By measuring gene expression, the probes are useful for predicting, diagnosing, grading, staging, monitoring and proposing diseases of the human heart and vascular system e.g. cardiovascular disease, hypertension, cardiac arrhythmias and congenital heart disease.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences).

Sequence 186 BP; 63 A; 34 C; 36 G; 53 T; 0 other;

Query Match	7.2%	Score 186;	DB 22;	Length 186;
Best Local Similarity	100.0%			

Best local similarity 100.0%; Pred. No. 5.8e-43;  
Matches 186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

253	GTTTTGACAACAATTCGTTTCAGGGCACCAGCAACAATATTTAGTCGAAAGTTCTTACCT	312
1	GTTTTGACAACAATTCGTTTCAGGGCACCAGCAACAATATTTAGTCGAAAGTTCTTACCT	60

313 TGTACAAATGATAAACAGATTGTATCTGCTCTGGAGATGGAGTAATATTTTATACCAAC 372

61 TGTACAAATGATAAACAGATTGTATCCTGCTCTGGAGATGGAGTAAATATTTATACCAAC 120

373 GTTGAGCAGATGCAGAAACCAACAGACAATGCCAATTACGTGTCATTATGGAACTACT 432

b  
121 GTTGAGCAAGATGCAGAAACCAACAGACAATGCCAATTTACGTGTCAATTATGGAAC TACT 180

433 TATGAG 438

b 181 TATCAG 186

Search completed: March 25, 2003, 23:52:13  
Job time : 632 secs

100

100

100

100

100

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2003, 23:21:03 ; Search time 4054 Seconds  
(without alignments)  
10306.945 Million cell updates/sec

Title: US-09-781-693a-1

Perfect score: 2580

Sequence: 1 atgtctcgggggtgctccta.....atgaataatgagatgagaa 2580

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 120 summaries

Database :

EST:

1: em\_estb:\*

2: em\_estb:\*

3: em\_estb:\*

4: em\_estb:\*

5: em\_estb:\*

6: em\_estb:\*

7: em\_estb:\*

8: em\_estb:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_est3:\*

12: gb\_est4:\*

13: gb\_est5:\*

14: gb\_est6:\*

15: em\_estb:\*

16: em\_estb:\*

17: gb\_gss:\*

18: em\_gss\_hum:\*

19: em\_gss\_inv:\*

20: em\_gss\_inv:\*

21: em\_gss\_vrt:\*

22: em\_gss\_vrt:\*

23: em\_gss\_mam:\*

24: em\_gss\_mus:\*

25: em\_gss\_mus:\*

26: em\_gss\_pro:\*

27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1981.4	76.8	2375	11 AK004618	
2	890.8	34.5	1045	14 BM807192	
3	853.6	33.1	1052	13 BM477718	
4	832.6	32.3	864	14 BQ433226	
5	827.4	32.1	853	14 BQ933148	
6	826.6	32.0	1074	14 BQ067602	

7	784.8	30.4	839	12	BG740812
8	781.4	30.3	805	12	BG686945
9	743	28.8	827	13	BI335444
10	735.8	28.5	951	12	BF980168
11	725.8	28.1	843	14	BQ229691
12	723	28.0	906	14	BQ433486
13	720.4	27.9	722	9	AL560929
14	716.8	27.8	791	13	BI834692
15	701.2	27.2	860	13	BI846254
16	684.2	26.5	809	13	BI464753
17	681.4	26.4	730	13	BI461216
18	678.8	26.3	729	12	BG682657
19	667.8	25.9	783	13	BI464045
20	665.8	25.8	754	12	BG718960
21	665.8	25.8	834	13	BI461627
22	663.6	25.7	850	13	BI464622
23	658.2	25.5	694	10	BE180558
24	658.8	25.4	739	12	BG117357
25	648.6	25.1	1075	14	BM923018
26	646.6	25.1	775	12	BE867206
27	641.8	24.9	806	12	BG772319
28	641.6	24.9	911	12	BG530857
29	640	24.8	648	14	BQ019842
30	634	24.6	634	12	BE747286
31	633.4	24.6	635	14	BM790042
32	631.4	24.5	707	12	BG719267
33	629.8	24.4	728	12	BG534181
34	628.6	24.4	869	13	BI459472
35	627.4	24.3	762	10	BE559905
36	627	24.3	717	13	BI559857
37	624.4	24.2	733	13	BI561174
38	624	24.2	1199	14	BM926145
39	623.6	24.2	666	10	BE180608
40	618.4	24.0	912	12	BF029089
41	616.6	23.9	709	12	BG720740
42	612.8	23.8	868	12	BE884542
43	611.6	23.7	918	12	BE165887
44	610.8	23.7	616	13	BI836271
45	605	23.4	624	10	BE180560
46	604	23.4	604	9	AL602747
47	601.6	23.3	735	12	BF672818
48	599.2	23.2	718	13	BI562377
49	596.8	23.1	730	12	BG724412
50	592.4	23.0	847	12	BG164080
51	588	22.8	597	14	BM726957
52	584.4	22.7	1141	12	BG255311
53	582.8	22.6	620	14	BG575260
54	580.6	22.5	922	12	BE882087
55	575.4	22.3	668	13	BM538035
56	574.8	22.3	593	9	AI832097
57	574	22.2	619	12	BF695020
58	573.2	22.2	676	12	BF979546
59	567.8	22.0	791	12	BG484572
60	566.6	22.0	956	12	BF662916
61	562.4	21.8	658	10	BE268392
62	557.4	21.6	578	10	BE180559
63	550	21.3	805	12	BG217514
64	547	21.2	547	9	AI927915
65	535	20.7	713	10	BE545703
66	511.6	19.8	550	10	AW992096
67	510.6	19.8	532	10	BE019647
68	509	19.7	832	14	BQ435888
69	508	19.7	508	10	BE504075
70	507.6	19.7	839	14	BQ221011
71	504.2	19.5	540	10	AW992159
72	504.2	19.5	675	12	BF671543
73	504	19.5	504	9	AA626631
74	498.2	19.3	1132	13	BM478778
75	496.2	19.2	567	9	AU146956
76	492.8	19.1	544	9	AU160319
77	490.8	19.0	510	10	BE019684
78	488	18.9	516	14	BM685550
79	487.8	18.9	711	12	BG567316

BG740812 602632794  
BG686945 602650339  
BI335444 602958334  
BF980168 602288183  
BQ229691 AGENCOURT  
BQ433486 AGENCOURT  
AL560929 AL560929  
BI834692 603090476  
BI462541 603208451  
BI464753 603202393  
BI461216 603206537  
BG682657 60268942  
BI464045 603202819  
BG718960 602699258  
BI461627 603203236  
BE180558 RC3-HT062  
BE117357 602346361  
BM923018 AGENCOURT  
BE867206 601442286  
BG772319 602722419  
BG530857 602559894  
BQ019842 UI-H-ED0-  
BE747286 601580236  
BM790042 K-EST0069  
BG719267 602690335  
BG534181 602553301  
BI459472 603200466  
BE559905 601345445  
BI559857 603252775  
BI561174 603253711  
BM926145 AGENCOURT  
BE180608 RC3-HT062  
BF029089 601764588  
BG720740 602691950  
BE884542 601508239  
BI65887 602344381  
BI836271 603083948  
BE180560 RC3-HT062  
AL602747 DFEZp686G  
BF672818 602152653  
BI562377 603255541  
BG724412 602696546  
BI64080 602341114  
BM726957 UI-E-EJ0-  
BG255311 602369894  
BQ575260 UI-H-EZ1-  
BE882087 601505375  
BM538035 ha89d10.g  
AI832097 tdl2f03.x  
BF695020 602083051  
BF979546 602288183  
BG484572 602505739  
BF662916 602144873  
BE268392 601125039  
BE180559 RC3-HT062  
RG217514 RST37224  
AI927915 wp03d06.x  
BE545703 601077439  
AW992096 CM3-BN002  
BE019647 bp27902.x  
BQ435888 AGENCOURT  
BE504075 h249d11.x  
BQ221011 AGENCOURT  
AW992159 CM3-BN002  
BF671543 602151475  
AA626631 ab51c10.r  
BM478778 AGENCOURT  
AU146956 AU146956  
AU160319 AU160319  
BE019684 db28902.x  
BM685550 UI-B-C10-  
BG567316 602585808

us-09-781-693a-1.rst

Wed Mar 26 09:38:52 2003

MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

20499374  
11042159

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kiyosawa, H., Kondo, S., Yamada, A., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujikawa, S., Inoue, K., Togawa, X., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
20530913  
11076861

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamada, A., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Quackenbush, J., Schriml, L. M., Stauber, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Balderelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Hill, D., Hofmann, C., Fujita, M., Gariboldi, M., Gustincich, S., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Marzari, M., Lee, N. H., Lyons, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S. and Hayashizaki, Y.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409 (6821), 685-690 (2001)  
21085660  
11217851

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

5 (bases 1 to 2975)  
Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arakawa, T., Balderelli, R., Bono, H., Brownstein, M., Bult, C., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hiraoka, T., Hori, F., Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J., Saito, H., Saito, R., Sakai, K., Sakai, K., Sano, H., Sasaki, D., Schriml, L., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yamada, A., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.  
Direct submission  
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp, URL: http://genome.gsc.riken.go.jp/; Tel: 81-45-503-9222, Fax: 81-45-503-9216)  
Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

Hayashizaki, Y.  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGACAGCGCGCACTCGAGTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' ...

COMMENT

W19888 zb37b10.e1  
AUI18130 AUI18130  
BE568492 601342582  
AL705201 DKFP6860  
BG101295 y69h10.y  
AUI44811 AUI44811  
AUI44811 DKFP686K  
AL705761 DKFP686K  
AW087744 x68f07.x  
BE397963 601290501  
AW741120 u28c05.y  
BG171187 602151026  
BI155467 602903850  
BF575729 602135352  
BF062454 7h59c11.x  
AW135604 UI-H-B11-  
BG719120 602699052  
BM539925 hb16a03.g  
AA47852 aa20a06.x  
BG775713 602650089  
AA479513 zv21f10.r  
BF354864 PM1-HT079  
BG750040 602708810  
BG041839 h10h02.y  
BE175440 RC5-HT058  
W06613 y71b09.r1  
NS4388 y723f08.sl  
BE532778 601233929  
BE718001 PM1-HT079  
AW119185 x279g08.x  
BE175441 RC5-HT058  
A1334344 qg31c08.x  
RI464683 603202313  
BB617380 BB617380  
BM634650 UI-E-C10-  
BG539579 602567522  
AA085991 zn84g02.r  
AW771580 hu58f02.x  
AL568781 AL568781  
BM538697 hb98g09.9  
BF561661 UI-R-C0-1  
AA447700 aa20a06.s

ALIGNMENTS

2975 bp mRNA linear HTC 19-JAN-2002  
Mus musculus adult male lung cDNA, RIKEN full-length enriched library, clone:120006M05; homolog to DJ703H14.1 PROTEIN (FRAGMENT), full insert sequence.

AK004618  
AK004618.1 GI:12835911  
HTC; CAP trapper.  
Mus musculus (strain:C57BL/6J) adult male lung cDNA to mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library  
clone:120006M05.

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
99279253  
10349636

2  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)

JOURNAL

**FEATURES**  
**source**

QY	421	TATGGA	ACTACTTATGAGATATGACTGTATGACCTTACCAATGACCTTTACACTTTTCTCTCTGT	480
DB	520	TATGGA	ACTACTTATGAGATATGACTGTATGACCTTACCAATGACCTTTACACTTTTCTCTCTGT	579
QY	481	GGTGA	GATGAACTGTATGAGTGGTGTGATACACGATCAAACTAGCTGCAACAAAGAA	540
DB	580	GGTGA	GATGAACTGTATGAGTGGTGTGATACACGATCAAACTAGCTGCAACAAAGAA	639
QY	541	GATTG	TAAAGATGATATTTAATTAAGTCTGACGCTGTGCGACCTCTGTGCTATTTGC	600
DB	640	GACTGT	TAAAGATGATATTTAATTAAGTCTGACGCTGTGCGACCTCTGTGCTATTTGC	699
QY	601	CCACG	AAATACCATATTTACCTTCTGCTGTGCTTCTGCTGACGCTCAGTACGAATATATGAT	660
DB	700	CCCCG	AGTACCATATTTACCTTCTGCTGTGCTTCTGCTGAGCTCAGTACGGAATGAT	759
QY	661	CGGCA	GAATGCTGGGCACAGAGCTACAGGGAATATGACGCTCAGGAGCTACTGTGAATG	720
DB	760	CGGCA	GAATGCTGGGCACAGAGCTACAGGGAATATGACGCTCAGGAGCTACTGTGAATG	819
QY	721	GTTCG	CCGCTTTATTCCTTCCATCTTAAATAAGTCTTCCGAGTGACATCTCTGTGT	780
DB	820	GTTCG	TCGATTTATACCTTCTCATCTTAGTAACAAATCATGCGAGTGACATCACTGTGT	879
QY	781	TACAG	TGAAGTGTCAAGAGATCTCGTTAGTTACTTTCAGATTACATATCTCTTTT	840
DB	880	TACAG	TGAAGTGTCAAGAGATCTTGTGAGTTATCTTCAGACTACATCTATCTTTT	939
QY	841	GACCG	AAAGATGATACAGCAGAGACTTAAACTCTTCTCGGGAAGAGAGAGAA	900
DB	940	GACCG	AAAGATGATGATGACAGAGAACTTAAACTCTTCTCGGGAAGAGAGAGAA	999
QY	901	GAGTG	CGGACACCCACAGTTAAGCGTTTGAGACTTCTGAGCTTCTGAGTACTGGA	960
DB	1000	GAGTG	ACAGAGCTTCAAGTGTGAGCTTCTGAGCTTCTGAGCTTCTGAGTACTGGA	1059
QY	961	CCGAG	ACGAGCGGAGAGTCAAGCAGAGACGAGATGAGAGAGAGTCCCAATGTGCA	1020
DB	1060	CCGAG	ACGAGCGGAGAGTCAAGCAGAGAGATGAGAGAGAGTCCCAATGTGCA	1119
QY	1021	TTGAT	GTCAGAGATGCTGATGATTTATCAAGTGGTTTGAAGACAGTGGAGTTGCA	1080
DB	1120	CTGAT	GTCAGAGATGCTGATGATTTATCAAGTGGTTTGAAGACAGTGGAGTTGCA	1179
QY	1081	CAAGC	AAATAGAGCAGGAGAGTCTCGACCCAGAGCTGGAACAAGTCAATCAGATAT	1140
DB	1180	CAAGC	AAATAGAGCAGGAGAGTCTCGACCCAGAGCTGGAACAAGTCAATCAGATAT	1239
QY	1141	TCAAC	TCTTCTCGGTCCTCATCAAGTCTGATTTGGAAGTGGTGAAGTCAAGTGGAA	1200
DB	1240	TCAAC	TCTTCTCGGTCCTCATCAAGTCTGATTTGGAAGTGGTGAAGTCAAGTGGAA	1299
QY	1201	GTAGAT	ACTCCAGTGAACAATTTCTTCAGCTTCTACATCTCTACATGTCAGCTCAG	1260
DB	1300	GTAGAT	ACTCCAGTGAACAATTTCTTCAGCTTCTACATCTCTACATGTCAGCTCAG	1356
QY	1261	GCTCAT	TGACATCATCTCCACAGAAAGCTCATCTTACTCTCTTGTATCTTCTTCCCA	1320
DB	1357	GCTCAT	TGACATCATCTCCACAGAAAGCTCATCTTACTCTCTTGTATCTTCTTCCCA	1416
QY	1321	GACAG	TGAACAAGGAGTCTGTTGAGGCTCTGACACACCAAGTGAACCTCAGTCTGAT	1380
DB	1417	GACAG	TGAACAAGGAGTCTGTTGAGGCTCTGACACCAATGCAATCATCATGTCAGAT	1476
QY	1381	ANACA	TATGAAAAGCTGAGCCCAACACAGGACAGGTGAACCTTTAAGTTTGCAC	1440
DB	1477	ANACA	TATGAGAGCTGAGCCCAACACAGGACAGGTGAACCTTTAAGTTTGCAC	1536
QY	1441	TACAG	CACAGAGGACACTTACAGCACATATAACTGAATTTACAGATGAATGGAGC	1500
DB	1537	TACAG	CACAGAGGACACTTACAGCACATATAACTGAATTTACAGATGAATGGAGC	1596

```

URES      Location/Qualifiers
          1..2975
             /organism="Mus musculus"
             /strain="C57BL/6J"
             /db_xref="FANTOM_DB:I200006M05"
             /db_xref="MGD:MGI:1907187"
             /db_xref="taxon:10090"
             /clone="I200006M05"
             /sex="male"
             /tissue_type="lung"
             /clone_lib="RIKEN full-length enriched mouse cDNA library"
             /dev_stage="adult"
          100..2730
             /note="data source:SPTK, source key:Q9UJD5, evidence:ISS homolog to DU703H14.1 PROTEIN (FRAGMENT) putative"
CDS

```

ATLVNHDGCVNTILNDWGTGEYLISGSDQDKLVISNPKYKVLTVISRGHREFTQIRUKLE  
LPCTDDQIVSGDGVIFVTEIQEADENRCQCTCHGTITTEIMVPNDPIYFLSC  
GEGTWRVDFTRIKTCKEDCKDILLNCRASIAVCAICPPVYLAAGCSDDSVRI  
YDRMLGTRATNGAGRTGTVARIPHSNKSRCVTSILCYSDGQEILVYSSDY  
IYVLPDKDTRBELTIPSAEERKEELRQPVKRLRLGDWDJTGPRAPSEREREDGE  
QSPNYSLQRMDSMLSRFWEASVAQNRGRGPRPGCTNPDVSTLPIYVPSPNL  
EYCTAMVDVNNALLOPSTSTDPVQAATAALESPRSSLSLSDPSEPRVSEAS  
GHAHQDNDNNEERLSPKPTGEPVILHYSTEGTWTSTLNFTDWSSTASSRGN  
GSHCKSEGEELCVPPSQVPEGDSETRAEPEELSEKGTLPENLTQDITAOQLNDP  
AEPDLSNGEKNFNSQSPGLPEGLTSETDRETCQASTASATRIASNKPELPST  
BAIEQATSEATRIHSANPELPSQTAIAPLHEDPDSQASDAQTDSDDDPVLPIK  
ARYTPGCDRSARVAQIEFFRRHKKERKEEMELDTNIRPLPKYKTKGHKNSMTIK  
EANFWAGNFVMSGDCGHIETWDRTHAELMLADHNVVNCLOPPFPDILASSGTD  
YDKIWPGLBESIRFNKLAERFDRTHAELMLADHNVVNCLOPPFPDILASSGTD  
LEGDRSEGGSQENEDDEE"

855 a 660 c 721 g 737 t 2 others

BASE COUNT  
ORIGIN

Query Match	76.8%;	Score 1981.4;	DB 11;	Length 2975;
Best Local Similarity	85.8%;	Pred. No. 0;	Mismatches	286;
Matches 2273;	Conservative	0;	Indels	90;
			Gaps	3;
QY	1	ATGTCTCGGGTCTCTACCCACACCTGTTGTGGAGCTGAGGAAAAGTCCCTCGGG	60	
Db	100	ATGGCTCGAGTGGCTCTCTGCCCGCACCTGTTGTGGAGCTGAGGAAAAGTCCCTCGGG	159	
QY	61	CTGGAGGACCGTCCCGCTCGGGAGTCGCTACCTCGGAGGAGAGAAATTTATCCAAAGA	120	
Db	160	CTGGAGGACCGTCCCGCTCGGGAGTCGCTACCTCGGAGGAGAGAAATTTATCCAAAGA	219	
QY	121	TAAACCTTGGAACCAACCTTAAATGTCATGATGGTTGTGTTAATACAACTCTGTTGGAAT	180	
Db	220	TAAACCTTGGAACCACTTAAATGTCATGATGGCTGTGTTAATACAACTCTGTTGGAAT	279	
QY	181	GACACTGAGAAATATATTTATCTGGCTCAGATGACACCAAAATTAGTAATAGTAATCCT	240	
Db	280	GACACTGAGAAATATATTTATCTGGCTCAGATGACACCAAAATTAGTAATAGTAATCCT	339	
QY	241	TACAGCAAAAGGTTTGTGACCAAAATTCGTTCAAGGCACCGAGCAACATATTTAGTGA	300	
Db	340	TACAGCAAAAGGTTTGTGACCAAAATTCGTTCAAGGCACCGAGCAACATATTTAGTGA	399	
QY	301	AAGTCTTACCTTGATCAATGATAAACAATGTTATCTCTCTCGAGATGGAGTAATA	360	
Db	400	AAGTCTTTCGCTGCACAGATGATAGCAGATGTGCTTGTCTCTCGAGATGGAGTCATA	459	
QY	361	TTTTATACCAAGCTTGACAAGATGCGAGAAACCAACAGACAAATGCCAATTTACGTGTCAT	420	
Db	460	TTTTATACCAATTTGACCAAGATGCGAGAACTACAGACACTGCCAATTTACGTGTCAT	519	

[illegible]





Wed Mar 26 09:38:52 2003

```

Db 721 CAGATGAATGGAGCAGTATAGCATCAAGTTCTAGAGGAATTTGGAGGCAATTCGAAATCTG 780
QY 1547 AGGGTCAGGAGGAATCTTTCGTCACAGAGCTCAGTCAACCA-CCAGAAGGAGACAGT 1605
Db 781 AGGGTCAGGAGGAATCTTTCGTCACAGAGCTCAGTCAACCA-CCAGAAGGAGACAGT 840
QY 1606 GAAACAAAAGCTCTCTGAAGATCATCAGAGGATGTG---ACAAAATATCAGGAAGAGTA 1662
Db 841 GAAACAAAAGCTCTCTGAAGATCATCAGAGGATGTGGAACAAAATATTCAGGAAGAGTA 900
QY 1663 TC-TGCAGAAAACCCAGTTGAGA---ACATATCATATATACACATC 1706
Db 901 TCTTGCAGAAAACCCAGTTGAGAAGCCCTTATCCATATATACCAATC 948

RESULT 4
BQ433226 864 bp mRNA linear EST 24-MAY-2002
LOCUS AGENCOURT_7774746 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:6023741
DEFINITION 5', mRNA sequence.
ACCESSION BQ433226
VERSION BQ433226.1 GI:21172302
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 864)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCPD/OTD/Gazdar
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13233 row: c column: 06
High quality sequence stop: 692.
Location/Qualifiers
1. .864
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6023741"
/clone_lib="NIH_MGC_68"
/tissue_type="large cell carcinoma"
/lab_host="PH10B (phage-resistant)"
/notes="Organ: lung; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.8 kb. Library constructed by Life
Technologies."
BASE COUNT 272 a 188 c 198 g 206 t
ORIGIN
Query Match 32.3%; Score 832.6; DB 14; Length 864;
Best Local Similarity 99.3%; Pred. NO. 1.6e-224;
Matches 857; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

QY 1686 CCATATCAATATACAAATCAGATAGTTTCACAGCAAGCATTGGATTCCAACTCAGG 1745
Db 1 CCATATCAATATACAAATCAGATAGTTTCACAGCAAGCATTGGATTCCAACTCAGG 60
QY 1746 AGAAGAATGACCTCAATCTTGTATCGCTCTGTGGGGTCCAGAGAATCTGTTTCATC 1805
Db 61 AGAAGAATGACCTCAATCTTGTATCGCTCTGTGGGGTCCAGAGAATCTGTTTCATC 120
QY 1806 TGAAAAAGCCAAAGAACCAAGAACTTCAGATCAGACTAGAGTGCTACCAATGA 1865
Db 121 TGAAAAAGCCAAAGAACCAAGAACTTCAGATCAGACTAGAGTGCTACCAATGA 180

```

```

QY 1866 AATAACACCAATCTCTGAGCTCAGTTTCCAAACAGAGCACCCTGGGCTTCAGCTCATGA 1925
Db 181 AATAACACCAATCTCTGAGCTCAGTTTCCAAACAGAGCACCCTGGGCTTCAGCTCATGA 240
QY 1926 AGAAACATCCACAGGAGCTCTGCTCTTCAGGACACAGATGACAGTATGATGATGATGATGAT 1985
Db 241 AGAAACATCCACAGGAGCTCTGCTCTTCAGGACACAGATGACAGTATGATGATGATGATGATGAT 300
QY 1986 CCTGATCCAGGTCGCAAGGTATCGAGCAGGACCTGGTGTATAGAGCGTCTGCTGTGTCGCCG 2045
Db 301 CCTGATCCAGGTCGCAAGGTATCGAGCAGGACCTGGTGTATAGAGCGTCTGCTGTGTCGCCG 360
QY 2046 TATTCAGGAGTCTTCAGACCGGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2105
Db 361 TATTCAGGAGTCTTCAGACCGGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 420
QY 2106 GAACATTAGAAGCGCGCTAGTAAATAATGTTTATAAGGCCATCGCAACTCCAGGACAAAT 2165
Db 421 GAACATTAGAAGCGCGCTAGTAAATAATGTTTATAAGGCCATCGCAACTCCAGGACAAAT 480
QY 2166 GATAAAGAAAGCCAAATTTCTGGGGTCTAACTTTGTAATGAGTGGTCTGACTGTGGCCA 2225
Db 481 GATAAAGAAAGCCAAATTTCTGGGGTCTAACTTTGTAATGAGTGGTCTGACTGTGGCCA 540
QY 2226 CATTTTCATCTGGATCGGCACACTGCTGAGCATTTGATGCTTCTGGAAGCTGATAATCA 2285
Db 541 CATTTTCATCTGGATCGGCACACTGCTGAGCATTTGATGCTTCTGGAAGCTGATAATCA 600
QY 2286 TGTGGTAAACTGCTCGACGCCACATCCGTTTGACCAATTTTAGCCTCATCTGGCATAGA 2345
Db 601 TGTGGTAAACTGCTCGACGCCACATCCGTTTGACCAATTTTAGCCTCATCTGGCATAGA 660
QY 2346 TTATGACATTAAGATCTGGTCAACATTAGAAGAGTCAAGGATTTTAAACCGAAACTTGC 2405
Db 661 TTATGACATTAAGATCTGGTCAACATTAGAAGAGTCAAGGATTTTAAACCGAAACTTGC 720
QY 2406 TGATGAAGTTTAACTCGAAACGAACCTCATGCTGGGAAGAACTAGAAACACCACTTACAGT 2465
Db 721 TGATGAAGTTTAACTCGAAACGAACCTCATGCTGGGAAGAACTAGAAACACCACTTACAGT 780
QY 2466 TCCAGCCTCTTTCATGTTGAGGAT-GTTGGCTTCACCTTAATCATATCCGAGCTGACCGGT 2524
Db 781 TCCAGCCTCTTTCATGTTGAGGATGTTGGGCTTCACCTTAATCATATCCGAGCTGACCGGT 840
QY 2525 TCGA-GGCTGACAGATCAGAAGG 2546
Db 841 TGGAGGGGTGACAGATCAGAAGG 863

RESULT 5
BQ933148 853 bp mRNA linear EST 21-AUG-2002
LOCUS AGENCOURT_8821472 Lupski_sciatic_nerve Homo sapiens cDNA clone
DEFINITION IMAGE:6203588 5', mRNA sequence.
ACCESSION BQ933148
VERSION BQ933148.1 GI:22348531
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 853)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:

```

Query Match	32.1%	Score	827.4	DB	14	Length	853
Best Local Similarity	98.7%	Pred.	No. 4.8e-223				
Matches	842	Conservative	0	Mismatches	10	Indels	1
QY	1249	ATGTCAGGCTCAGGCTCATTCGACATCATCTCCACAGAAAGCCCTCATCTACTCCTTTG	1308				
Db	1	AUTGTCAGGCTCATTCGACATCATCTCCACAGAAAGCCCTCATCTACTCCTTTG	60				
QY	1309	CTATCTTCCAGACAGTGAACAAAGGCGAGTCTGTGTGAGCATCTGGACACCCACACAT	1368				
Db	61	CTATCTTCCAGATAGTGAACAAAGGCGAGTCTGTGTGAGCATCTGGACACCCACACAT	120				
QY	1369	CATCAGCTCGATAACAATATGAAAAGCTGAGCCCCAACCCAGGACAGGTGAACCAAGTT	1428				
Db	121	CATCAGCTCGATNACATATGAAAAGCTGAGCCCCAACCCAGGACAGGTGAACCAAGTT	180				
QY	1429	TTAAGTTTGCCTACAGCAGACAGGAACAACCTACAAGCACATAAACTGAACTTTACA	1488				
Db	181	TTAAGTTTGCACCTACAGCAGAGGAACAACCTACAAGCACATAAACTGAACTTTACA	240				
QY	1489	GATGAATGAGCAGTAGTACATCAACTCTTAGAGGAATTTGGGAGCCATTCGCAATCTGAG	1548				
Db	241	GATGAATGAGCAGTAGTACATCAAGTCTTAGAGGAATTTGGGAGCCATTCGCAATCTGAG	300				
QY	1549	GGTCAGGAGGAATCTTTTCGTCACAGAGCTCAGTGCACACCCAGCAAGGAGAGACAGTGAA	1608				
Db	301	GGTCAGGAGGAATCTTTTCGTCACAGAGCTCAGTGCACACCCAGCAAGGAGAGACAGTGAA	360				
QY	1609	ACAAAAGCTCCTGAGAATCATCAGAGGATGTGCACAAAATATCAGGAAGGAGTATTCGA	1668				
Db	361	ACAAAAGCTCCTGAGAATCATCAGAGGATGTGCACAAAATATCAGGAAGGAGTATTCGA	420				
QY	1669	GAAAACCCAGTTGAGAACCATATCAATATAACACAAATCAGATAAGTTTCACAGCCAAAGCCA	1728				
Db	421	GAAAACCCAGTTGAGAACCATATCAATATAACACAAATCAGATAAGTTTCACAGCCAAAGCCA	480				
QY	1729	TTGGATTCCAACCTCAGGAGAAAGAAATGACCTCAATCTTGATCGCTCTGTGGGGTTCCA	1788				
Db	481	TTGGATTCCAACCTCAGGAGAAAGAAATGACCTCAATCTTGATCGCTCTGTGGGGTTCCA	540				
QY	1789	GAAGAATCTGCTTCATCTCAAAAAGCCAGGAAACCTTCAGATCAGACTAGCACT	1848				
Db	541	GAAGAATCTGCTTCATCTGAANAAGCCAGGAAACCTTCAGATCAGACTAGCACT	600				
QY	1849	GAGAGTGCTACCAATGAAATTAACCAATCTGAGCCCTCAGTTCCAAAACAGAAAGCCACT	1908				
Db	601	GAGAGTGCTACCAATGAAATTAACCAATCTGAGCCCTCAGTTCCAAAACAGAAAGCCACT	660				



QY 1258 CAGGCTCATTCGACATCATCTCCACAGAAAGCCCTCATCTACTCTTGTCTATCTTCT 1317  
 Db 601 CAGGCTCATTCGACATCATCTCCACAGAAAGCCCTCATCTACTCTTGTCTATCTTCT 660

QY 1318 CCAGACAGTGAACAAAGGAGCTGTGTGGAGCATCTGGACACACACATCATCATCTCT 1377  
 Db 661 CCAGATAGTGAACAAAGGAGCTGTGTGGAGCATCTGGACACACACATCATCATCTCT 720

QY 1378 GATACAAATATGAAGAAGCTGAGCCCAACAGGACAGGAGTGAACAGCTTTTAAGTTTG 1437  
 Db 721 GATACAAATATGAAGAAGCTGAGCCCAACAGGACAGGAGTGAACAGCTTTTAAGTTTG 780

QY 1438 CACTACACACACA-GAAGACACAACTACACAGACACATAAATGAAGTGAAGTT 1484  
 Db 781 CACTACACACAGGAAGAACCACTACACAGACACATAAATGAAGTGAAGTT 828

RESULT 8  
 BG686945

LOCUS BG686945 805 bp mRNA linear EST 01-MAY-2001

DEFINITION 602650939F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4763336 5', mRNA sequence.

ACCESSION BG686945

VERSION BG686945.1 GI:13918342

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 805)

AUTHORS NTH-MGC http://mgc.nci.nih.gov/

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
 Plate: LICM1618 row: n column: 09  
 High quality sequence stop: 805.  
 Location/Qualifiers  
 1..805  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4763336"  
 /clone\_lib="NIH\_MGC\_48"  
 /issue\_type="primary B-cells from tonsils (cell line)"  
 /lab\_host="Dn10B (phage-resistant)"  
 /note="Organ: B-cells; Vector: pOTB7; Site: 1: XhoI; Site: 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: This is a NIH\_MGC Library."  
 274 a 179 c 178 g 174 t

BASE COUNT 274 a 179 c 178 g 174 t

ORIGIN

Query Match 30.3%; Score 781.4; DB 12; Length 805;  
 Best Local Similarity 99.1%; Pred. No. 5.2e-210;  
 Matches 796; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

QY 985 CGAAGACGAGATGGAGACGAGTCCCAATGTCTCATGTATGACAGAAATGCTGATG 1044  
 Db 2 CGAAGACGAGATGGAGACGAGTCCCAATGTCTCATGTATGACAGAAATGCTGATG 61

QY 1045 TTATCAAGATGGTTTGAAGAACGAAAGTGGTTCACAAAGCAATACAGGACGAGGAAGA 1104

Db 62 TTATCAAGATGGTTTGAAGAACGAAAGTGGTTCACAAAGCAATAGAGGAGGAGGAAGA 121

QY 1105 TCTCGACCCAGAGGTGGAAACAGTCAATCAGATATTTCAACTCTTCTCCTACGGTCCCATCA 1164

Db 122 TCTCGACCCAGAGGTGGAAACAGTCAATCAGATATTTCAACTCTTCTCCTACGGTCCCATCA 181

QY 1165 AGTCTCATTTGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1224

Db 182 AGTCTCATTTGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 241

QY 1225 CTTGAGCCCTTCTACATCTCTTACATCTCTTACATCTCTTACATCTCTTACATCTCTTACATCTCT 1284

Db 242 CTTGAGCCCTTCTACATCTCTTACATCTCTTACATCTCTTACATCTCTTACATCTCTTACATCTCT 301

QY 1285 GAAAGCCCTTCTACATCTCTTACATCTCTTACATCTCTTACATCTCTTACATCTCTTACATCTCT 1344

Db 302 GAAAGCCCTTCTACATCTCTTACATCTCTTACATCTCTTACATCTCTTACATCTCTTACATCTCT 361

QY 1345 GAGCATCTTGACACACACACATCATCATCTCTTACATCTCTTACATCTCTTACATCTCTTACATCTCT 1404

Db 362 GAGCATCTTGACACACACACATCATCATCTCTTACATCTCTTACATCTCTTACATCTCTTACATCTCT 421

QY 1405 AAACAGGGACAGGTGAACCACTTTTAAAGTTTGCACACTACAGCAGAGGAAAGTGAAGTGAAGTGAAGT 1464

Db 422 AAACAGGGACAGGTGAACCACTTTTAAAGTTTGCACACTACAGCAGAGGAAAGTGAAGTGAAGTGAAGT 481

QY 1465 AGCACAATAAATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1524

Db 482 AGCACAATAAATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 541

QY 1525 ATTGGGAGCCATTGCAAACTCTGAGGCTCAGGAGGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1584

Db 542 ATTGGGAGCCATTGCAAACTCTGAGGCTCAGGAGGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 601

QY 1585 CAACACACAGAGGAGGAGACAGTGAACAAAGTCTCTCTGAAGAAATCATCAGAGGATGTGACA 1644

Db 602 CAACACACAGAGGAGGAGACAGTGAACAAAGTCTCTCTGAAGAAATCATCAGAGGATGTGACA 661

QY 1645 AAATATCAGGAAGGAGTATCTCGAAGAAACCCAGTGTGAGAACCATATCAATATACACAA 1704

Db 662 AAATATCAGGAAGGAGTATCTCGAAGAAACCCAGTGTGAGAACCATATCAATATACACAA 721

QY 1705 TCAGATAAGTTTACAGCCAGCCATT-GGATTTCAACTCAGGAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1763

Db 722 TCAGATAAGTTTACAGCCAGCCATTGGGATTTCAACTCAGGAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 781

QY 1764 TCTTGATCGCTCTTGTGGGGTTC 1786

Db 782 TCTTGATCGCTCTTGTGGGGTTC 804

RESULT 9

LOCUS BI335444

DEFINITION 602998334F1 NIH\_MGC\_12 Homo sapiens cDNA clone IMAGE:5140544 5', mRNA sequence.

ACCESSION BI335444

VERSION BI335444.1 GI:15020101

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 827)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: Incyte Genomics, Inc.

DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM11346 row: c column: 09  
High quality sequence stop: 754.  
Location/Qualifiers

FEATURES  
source

1. .827  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5140544"  
/clone\_lib="NIH\_MGC\_12"  
/tissue\_type="cervical carcinoma cell line"  
/lab\_host="DH10B"  
/note="Organ: cervix; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dt.  
Average insert size 1.4 kb. Library prepared by Life  
Technologies."  
BASE COUNT 253 a 160 c 183 g 231 t  
ORIGIN  
Query Match 28.8%; Score 743; DB 13; Length 827;  
Best Local Similarity 97.68; Pred. No. 4.1e-199;  
Matches 807; Conservative 0; Mismatches 15; Indels 5; Gaps 5;  
QY 45 GAAAGGTCCTCGGCTGGAGGACCCGCTCCCGGCTCGGAGTCGCTACCTGGGAAGAAG 104  
DB 1 GAAAGGTCCTCGGCTGGAGGACCCGCTCCCGGCTCGGAGTCGCTACCTGGGAAGAAG 60  
QY 105 AGAATTTCAAGATTAATAAATTTGAAGCAACCTTAAATGTCATGATGTCGTCTTAA 164  
DB 61 AGAATTTCAAGATTAATAAATTTGAAGCAACCTTAAATGTCATGATGTCGTCTTAA 120  
QY 165 TACAATCTGTTGGAATGACACTGGAGATATATTTATCTGTGCTCAGATGACACCAAAAT 224  
DB 121 TACAATCTGTTGGAATGACACTGGAGATATATTTATCTGTGCTCAGATGACACCAAAAT 180  
QY 225 AGTA-ATPAGTAATCTTACAGCAGAAAGTTTGAACAACAAATTCGTCAGGGCACCGAG 283  
DB 181 AGTAGAGGAGTAATCTTACAGCAGAAAGTTTGAACAACAAATTCGTCAGGGCACCGAG 240  
QY 284 CAAACATATTTAGTCCAAAGTTCTTACCTTGTACAAATGATAACAGATGTCCTGCT 343  
DB 241 CAAACATATTTAGTCCAAAGTTCTTACCTTGTACAAATGATAACAGATGTCCTGCT 300  
QY 344 CTGGAGATGGAGTAATATTTATACCAACGTTGACAAAGATGACAGAAACCAACAGACAAAT 403  
DB 301 CTGGAGATGGAGTAATATTTATACCAACGTTGACAAAGATGACAGAAACCAACAGACAAAT 360  
QY 404 GCCAATTTACGTCATTTATGGAACACTTATGAGATATGACTGTACCAATGACCCCTT 463  
DB 361 GCCAATTTACGTCATTTATGGAACACTTATGAGATATGACTGTACCAATGACCCCTT 420  
QY 464 ACACATTTCTCTCTGTTGTTGAGTGAAGTGAACCTGTAGGTGTTGATACACGCATCAAAA 523  
DB 421 ACACATTTCTCTCTGTTGTTGAGTGAAGTGAACCTGTAGGTGTTGATACACGCATCAAAA 480  
QY 524 CTAGCTGCACAAAAGAGATTTAAAGATGATATTTTAAATTAAGTGTGACAGCT 583  
DB 481 CTAGCTGCACAAAAGAGATTTAAAGATGATATTTTAAATTAAGTGTGACAGCT 540  
QY 584 CGTCTGTTGCTATTTGCCCAACCAATACCATATTTACCTGCTGTTGTTGTTCTGACAGCT 643  
DB 541 CGTCTGTTGCTATTTGCCCAACCAATACCATATTTACCTGCTGTTGTTGTTCTGACAGCT 600  
QY 644 CAGTAGGAA-TATATGATCGGGGAATGCTGGGCACAAAGAGCTACAGGGGAATTTATGAGCT 702  
DB 601 CAGTAGGAACTATATGATCGGGGAATGCTGGGCACAAAGAGCTACAGGGGAATTTATGAGCT 660  
QY 703 CGAGGAGCTACTGGAATGTTGCC-CTTTTATCTCTCCCATCTTAAT-AATAAGTCTCT 760  
DB 661 CGAGGAGCTACTGGAATGTTGCCCGGTTTAAATTCCTTCCCATCTTAATTAAGTCTCT 720

QY 761 GCAGAGTGACATCTCTGTTTACAGTGAAGATGTCAG-AGATTCCTGTTAGTACTCT 819  
DB 721 GCAGAGTGACATCTCTGTTTACAGTGAAGATGTCAG-AGATTCCTGTTAGTACTCT 780  
QY 820 TCAGATTACATATATCTTTTGGACCCGAAAGATGATACAGCAGCAGAGA 866  
DB 781 TCAGATTACATATATCTTTTGAACCCGAAAGATGATACAGCAGCAGAGA 827

RESULT 10

BF980168/c 951 bp mRNA linear EST 23-JAN-2001  
LOCUS 602288183T1 NIH\_MGC\_97 Homo sapiens CDNA clone IMAGE:4373781 3',  
DEFINITION mRNA sequence.

ACCESSION BF980168

VERSION BF980168.1 GI:12347383

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 951)  
NIH-MGC http://mgi.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cspas-remail.nih.gov

REFERENCE

1 (bases 1 to 951)  
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki  
Toshiyuki and Piero Carninci (RIKEN)  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM10036 row: f column: 22  
High quality sequence start: 22  
High quality sequence stop: 792.

FEATURES

source

1. .951

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4373781"

/clone\_lib="NIH\_MGC\_97"

/lab\_host="DH10B"

/note="Organ: testis; Vector: pBluescriptR (modified  
pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI (gtcagag  
); Oligo-dr primed using primer 5'-TTTTTTTTTTTTTTVN-3',  
size-selected for average insert size 2.2 kb and  
normalized to ROT 5. This is a primary library enriched  
for full-length clones and constructed using the  
Cap-trapper method (Carninci, in preparation). Library  
constructed by M. Brownstein (NIH/NHGRI, National  
Institutes of Health). Note: this is a NIH\_MGC Library."  
204 a 234 c 225 g 288 t

Query Match 28.5%; Score 735.8; DB 12; Length 951;

Best Local Similarity 95.8%; Pred. No. 4.8e-197;

Matches 821; Conservative 0; Mismatches 27; Indels 9; Gaps 6;

QY 1696 ATAACACATCATGATAAGTTCACAGCCAGCCATTGGATTCCCACTCAGAGAGAAAT 1755

DB 850 AAACACACCCGAGAGGTTTCACAGCCAGCCC--TGGATCCAACCTCAGAG-AGAAGT 794

QY 1756 GACCTCAATCTTGATGCTCTTGTGGGGTTCAGAGAAATCTGCTTCATCTGAAAAGCC 1815

DB 793 GACCTCAATCT--GATCGCTCTGTGGGTCCTCAGAGAAATCTGCTTCATCTGAAAAGCC 737

QY 1816 AAGGAACCAAGAACT--TCAGATCAGACTAGCAGTGTGTACCAATGAAAATAACAC 1874

DB 736 CAGGAACCAAGAACTGTACATCAGACTAGCAGGAGTGCTACCAATGAAAAT-ACAC 678





```

QY 2233 ATC 2235
Db 792 TTC 794

RESULT 12
LOCUS B0433486
DEFINITION AGENCOURT_7805625 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6067882
5', mRNA sequence.
ACCESSION BQ433486
VERSION BQ433486.1 GI:21172562
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 906)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: c9abps-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LAM13348 row: b column: 11
High quality sequence stop: 451.
Location/Qualifiers
1. 906
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6067882"
/clone_lib="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: testis; Vector: pCMV-SPORT6; Site1: NotI;
Site2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 291 a 216 c 209 g 190 t
ORIGIN
Query Match 28.0%; Score 723; DB 14; Length 906;
Best Local Similarity 96.3%; Pred. No. 2e-193;
Matches 773; Conservative 0; Mismatches 25; Indels 5; Gaps 3;

QY 970 AGCCCGGAGAGTGAACGACGAGATGGAGAGCAGAGTCCCAATGTGTCAATGATGACG 1029
Db 1 AGCCCGGAGAGTGAACGACGAGATGGAGAGCAGAGTCCCAATGTGTCAATGATGACG 60

QY 1030 AGAATGCTGTATGTATCAAGATGGTTTGAAGAAGCAAGTGGAGTTGCACAAAGCAAT 1089
Db 61 AGAATGCTGTATGTATCAAGATGGTTTGAAGAAGCAAGTGGAGTTGCACAAAGCAAT 120

QY 1090 AGAGGACGAGGAAGATCTGACCCAGGAGTGGACAAAGTCAATCAGATATTTCAACTCTT 1149
Db 121 AGAGGACGAGGAAGATCTGACCCAGGAGTGGACAAAGTCAATCAGATATTTCAACTCTT 180

QY 1150 CCTACGGTCCCATCAAGTCCCTGATTTGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1209
Db 181 CCTACGGTCCCATCAAGTCCCTGATTTGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 240

QY 1210 CCAGCTGAACAATTTCTTCAGCCCTTCACATCTCTACATGTCAGCTCAGGCTCATTCG 1269
Db 241 CCAGCTGAACAATTTCTTCAGCCCTTCACATCTCTACATGTCAGCTCAGGCTCATTCG 300

```

```

QY 1270 ACATCATCTCCACACAGAAAGCCCTCATTTCTACTCCTTTTGGTATCTTCTCCAGACAGTGAA 1329
Db 301 ACATCATCTCCACACAGAAAGCCCTCATTTCTACTCCTTTTGGTATCTTCTCCAGACAGTGAA 360

QY 1330 CAAAGGCAGTCTGTTGAGGCATCTGGACACACACACATCATCATGCTGTGATAACAATAAT 1389
Db 361 CAAAGGCAGTCTGTTGAGGCATCTGGACACACACACATCATCATGCTGTGATAACAATAAT 420

QY 1390 GAAAGCTGAGCCCAAAA--CCAGGACACAGGTGAACACAGTCTTTTAAGTTTGCACACTACAGCA 1447
Db 421 GAAAGCTGAGCCCAAAA--CCAGGACACAGGTGAACACAGTCTTTTAAGTTTGCACACTACAGCA 480

QY 1448 CAGAAGGAACAACACTACAAGCACAATAAACTGAACCTTTACAGATGAATGAGCAGATATAG 1507
Db 481 CAGAAGGAACAACACTACAAGCACAATAAACTGAACCTTTACAGATGAATGAGCAGATATAG 540

QY 1508 CATCAAGTTCTAGAGAAATTTGGGAGCCATTGCAAAATCTGAGGCTCAGGAGGAATCTTTTCG 1567
Db 541 CATCAAGTTCTAGAGAAATTTGGGAGCCATTGCAAAATCTGAGGCTCAGGAGGAATCTTTTCG 600

QY 1568 TCCACAGAGCTCAGTGCACACACACAGGAGGAGAGTATCTGCAGAAACCCAGTTGAGAAC 1627
Db 601 TCCACAGAGCTCAGTGCACACACACAGGAGGAGAGTATCTGCAGAAACCCAGTTGAGAAC 660

QY 1628 CATCAGAGGATGTGACAAAATATCAGGAAGGAGTATCTGCAGAAACCCAGTTGAGAAC 1687
Db 661 CATCAGAGGATGTGACAAAATATCAGGAAGGAGTATCTGCAGAAACCCAGTTGAGAAC 720

QY 1688 ATATCAATAAACAACATCAGATAAGTT--CACAGCCCAAGCATTGGATTCCCAAC--TCAG 1744
Db 721 CTTTTCAGTATACCCCATCGGATAAGTTCCCGAGCAAGCCATTGGATTCCCAACCTCCG 780

QY 1745 GAGAAAGAAATGACCTCAATCTT 1767
Db 781 AAAAAAGAAATGACCTCAATCTT 803

RESULT 13
LOCUS AL560929
DEFINITION AL560929 LTI_NFL010_BC2 722 bp mRNA linear EST 16-FEB-2001
prime, mRNA sequence.
VERSION AL560929
AL560929.1 GI:12907861
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 722)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1. 722
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DL005Y113"
/clone_lib="LTI_NFL010_BC2"
/sex="male"
/tissue_type="B cells from Burkitt lymphoma"
/notes="Vector: pCMVSPORT6; Site1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371

```



---

```

Db 481 CAATCTGAGCCTCAGTTCCTCAACAGAGCCACTGGGCTTCAGTTCATGAAGAACATC 540
QY 1935 CACCAGGACTCTGCTCTTCCAGGACACAGATGACAGTATGATGACCCAGTCTGATCCC 1994
Db 541 CACCAGGACTCTGCTCTTCCAGGACACAGATGACAGTATGATGACCCAGTCTGATCCC 600
QY 1995 AGGTGCAAGGTATCGAGCAGGACCTGGTGATAGACGCTCTGCTGTTGCCGGTATTCAGGA 2054
Db 601 AGGTGCAAGGTATCGAGCAGGACCTGGTGATAGACGCTCTGCTGTTGCCGGTATTCAGGA 660
QY 2055 GTTCTTCAGACGGGAGAAAGAAAGAAAGAA-----TGGAGAATTTGGATCTTTGAACA 2110.
Db 661 GTTCTTCAGACGGGACACAGACAGGACAGAAATGGCAGCAATTTGGGATCTTTGAACA 720
QY 2111 -TTAGAAGCGCGTATGATAAATGTTTATAAAGCCATCGCACTCCAGGACAATGATA 2169
Db 721 TTTAGAAGCGCGTATGATAAATGTTTATAAAGCCATCGCACTCCAGGACAATGATA 780
QY 2170 AA 2171
Db 781 AA 782

RESULT 15
BI462541
LOCUS 603205451F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5271157 5',
DEFINITION mRNA sequence.
ACCESSION BI462541
VERSION BI462541.1 GI:152533197
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 860)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLNL1684 row: e column: 14
High quality sequence stop: 774.
Location/Qualifiers
1. 860
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5271157"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/notes="Organ: testis; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTNN-3',
size-selected for average insert size 2.2 kb and
normalized to 50x. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
instituted by M. Brownstein (NHGRI), National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 254 a 165 c 200 g 241 t
ORIGIN

Query Match 27.2%; Score 701.2; DB 13; Length 860;
Best Local Similarity 95.9%; Pred. No. 3e-187;
Matches 752; Conservative 0; Mismatches 28; Indels 4; Gaps 3;

```

```

QY 220 AAATTAAGTAATTAATCCTTTACAGCAGAAAGGTTTGTGACAAATTCGTTTCAGGGCAC 279
Db 34 AAATTAATTTTAAAGGAGAGGTTTCTCATCAGAGGTTTGTGACAAATTCGTTTCAGGGCAC 93
QY 280 CGAGCAACATATTTAGTGTCAAGGTTTCTTACCTTGTCAAAATGATTAACAGATTGTATCC 339
Db 94 CGAGCAACATATTTAGTGTCAAGGTTTCTTACCTTGTCAAAATGATTAACAGATTGTATCC 153
QY 340 TGCTCTGGAGATGAGTAATATTTTATACCAACGTTGAGCAAGATGAGCAAAACCAACAGA 399
Db 154 TGCTCTGGAGATGAGTAATATTTTATACCAACGTTGAGCAAGATGAGCAAAACCAACAGA 213
QY 400 CAATGCCAATTTAGCTGTCAATTTATGGAACACTTATGAGATTTATGCTGTACCAATGAC 459
Db 214 CAATGCCAATTTAGCTGTCAATTTATGGAACACTTATGAGATTTATGCTGTACCAATGAC 273
QY 460 CCTTACACTTTTCTCTCTTGTGGTGAAGATGGAATCTGTTAGTGTGTTGATACAGCATC 519
Db 274 CCTTACACTTTTCTCTCTTGTGGTGAAGATGGAATCTGTTAGTGTGTTGATACAGCATC 333
QY 520 AAAAATAGCTGCACAAAAGAAAGATTGTAAGATGATATTTTAAATTAACCTGTCGAGTCT 579
Db 334 AAAAATAGCTGCACAAAAGAAAGATTGTAAGATGATATTTTAAATTAACCTGTCGAGTCT 393
QY 580 GCCAGCTCTGTTGCTATTTTGGCCCAATATACATATTTACCTTGTGTTGTTGTTCTGAC 639
Db 394 GCCAGCTCTGTTGCTATTTTGGCCCAATATACATATTTACCTTGTGTTGTTGTTCTGAC 453
QY 640 AGCTCAGTACGAATATATGATCGGCGAATGCTGGGCAACAGAGCTACAGGGAATTTATGCA 699
Db 454 AGCTCAGTACGAATATATGATCGGCGAATGCTGGGCAACAGAGCTACAGGGAATTTATGCA 513
QY 700 GGTGAGGAGTACTGGAATGTTGCCCGTGTATTCCTTCCCATCTTAATAATTAAGTCC 759
Db 514 GGTGAGGAGTACTGGAATGTTGCCCGTGTATTCCTTCCCATCTTAATAATTAAGTCC 573
QY 760 TGCAGAGTGACATCTCTGTTTACAGTGAAGATGTCAGAGATTCGTTAGTTACTCT 819
Db 574 TGCAGAGTGACATCTCTGTTTACAGTGAAGATGTCAGAGATTCGTTAGTTACTCT 633
QY 820 TCAGATTACATATATCTTTTGAACCCGAGATGATACAGCAGAGAACTTTAAACTCT 879
Db 634 TCAGATTACATATATCTTTTGAACCCGAGATGATACAGCAGAGAACTTTAAACTCT 693
QY 880 TCTGCGGAAGAGAGAGAGAGAGAGTGGCCACACACAGTTAAAGCGTTTGAGACTCT 939
Db 694 TCTGCGGAAGAGAGAGAGAGAGTGGCCACACACAGTTAAAGCGTTTGAGACTCT 751
QY 940 GGTGATTGGTCAATCTGACCTGACCCAGAGCAAGCGGAGAGTGAACGAGAGATGGA 999
Db 752 GGGGA-TGGTCAATCTGACCCAGAGCAAGCGGAGAGTGAACGAGAGATGGA 809
QY 1000 GAGC 1003
Db 810 AACC 813

RESULT 16
BI464753
LOCUS 603202393F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5268215 5',
DEFINITION mRNA sequence.
ACCESSION BI464753
VERSION BI464753.1 GI:15255409
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 809)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)

```

JOURNAL  
COMMENT

Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: c9apbs-r@mail.nih.gov  
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM11676 row: j column: 24  
High quality sequence stop: 761.  
Location/Qualifiers  
1..809

FEATURES  
source

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5268215"  
/clone.lib="NIH\_MGC\_97"  
/lab\_host="DH10B"  
/note="Organ: testis; Vector: pBluescriptR (modified pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI (gtcgag); Oligo-dr primed using primer 5'-TTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.2 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH\_MGC Library."  
246 a 149 c 188 g 226 t

BASE COUNT 246 a 149 c 188 g 226 t  
ORIGIN  
Query Match 26.5%; Score 684.2; DB 13; Length 809;  
Best Local Similarity 93.6%; Pred. No. 1.9e-182;  
Matches 735; Conservative 0; Mismatches 48; Indels 2; Gaps 2;  
QY 210 AGATGACACCAATAGTAATTAATCCCTACAGCAAGAAAGTTTGCACCAACATTCG 269  
DB 23 AGAGCACTGAGAAATACCTTTTAAGGAGAGGTTCTCATCAGAGTTTGCACCAACATTCG 82  
QY 270 TTCAGGCGACCGAGCAACATATTAGTGCAGAAAGTTCTTACCTGTGACAAATGATAACA 329  
DB 83 TTCAGGCGACCGAGCAACATATTAGTGCAGAAAGTTCTTACCTGTGACAAATGATAACA 142  
QY 330 GATTGTATCCTGCTCGAGATGAGTAATATTATACCAACGTTGAGCAAGATGAGA 389  
DB 143 GATTGTATCCTGCTCGAGATGAGTAATATTATACCAACGTTGAGCAAGATGAGA 202  
QY 390 AACCAACAGACAATGCCAATTTAGCTGTCAATTATGGAACACTTATGAGATTATGACTGT 449  
DB 203 AACCAACAGACAATGCCAATTTAGCTGTCAATTATGGAACACTTATGAGATTATGACTGT 262  
QY 450 ACCCAATGACCTTACACTTTCTCTCTGTTGTTGAGATGGAACACTGTTAGTGGTTGA 509  
DB 263 ACCCAATGACCTTACACTTTCTCTCTGTTGTTGAGATGGAACACTGTTAGTGGTTGA 322  
QY 510 TACACGATCAAACTAGCTGCACAAAGAGATGTAAGATGATATTTTAATTAACCTG 569  
DB 323 TACACGATCAAACTAGCTGCACAAAGAGATGTAAGATGATATTTTAATTAACCTG 382  
QY 570 TCAGCTGCTGCCACGCTCTGCTTAFTTCCCAACCAATACCATATACCTTGTGCTGTTGG 629  
DB 383 TCAGCTGCTGCCACGCTCTGCTTAFTTCCCAACCAATACCATATACCTTGTGCTGTTGG 442  
QY 630 TTGTTCTGACAGCTCAGTACGATATATGATCGCGAATCCTGGGCAACAGAGTACAGG 689  
DB 443 TTGTTCTGACAGCTCAGTACGATATATGATCGCGAATCCTGGGCAACAGAGTACAGG 502  
QY 690 GAATTATGACAGCTGAGGGACTACTGGAATGTTGCCGCTTTTATCTCTCCATCTTAA 749  
DB 503 GAATTATGACAGCTGAGGGACTACTGGAATGTTGCCGCTTTTATCTCTCCATCTTAA 562  
QY 750 TAATAAGTCTCGAGATGACATCTCTGTGTTTACAGTGAAGATGGTCAAGAGATTCCTCGT 809

Db 563 TAATAAGTCTCGAGAGTACATCTCTGTTTACAGTGAAGATGGTCAAGAGATTCCTCGT 622  
QY 810 TAGTTACTCTTCAGATTACATATATCTTTTGTACCCGAAAGATGATACAGCAGAGAAC- 868  
DB 623 TAGTTACTCTTCAGATTACATATATCTTTTGTACCCGAAAGATGATACAGCAGAGAAC- 682  
QY 869 TTAATAACTCTCTCTCGGAG- 928  
DB 683 TTAATAACTCTCTCTCGGAG- 741  
QY 929 TGAGACTCTCTGTTGATGTTGTCAGATCTGGACCCAGAGCAAGCGCGGAGAGTGAACGAG 988  
DB 742 TGAGACTCTCTGTTGATGTTGTCAGATCTGGACCCAGAGAGAGAGAGAGAGAGAGAGAG 801  
QY 989 AACGA 993  
DB 802 AACGA 806  
RESULT 17  
BI461216  
LOCUS 603206537F1 NIH\_MGC\_97 Homo sapiens cDNA clone IMAGE:5272400 5',  
DEFINITION mRNA sequence.  
ACCESSION BI461216  
VERSION BI461216.1 GI:15251872  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 730)  
NIH-MGC http://mgc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: c9apbs-r@mail.nih.gov  
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM11687 row: i column: 09  
High quality sequence stop: 723.  
Location/Qualifiers  
1..730  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5272400"  
/clone.lib="NIH\_MGC\_97"  
/lab\_host="DH10B"  
/note="Organ: testis; Vector: pBluescriptR (modified pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI (gtcgag); Oligo-dr primed using primer 5'-TTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.2 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH\_MGC Library."  
221 a 137 c 165 g 207 t  
BASE COUNT 221 a 137 c 165 g 207 t  
ORIGIN  
Query Match 26.4%; Score 681.4; DB 13; Length 730;  
Best Local Similarity 97.7%; Pred. No. 1.1e-181;  
Matches 691; Conservative 0; Mismatches 16; Indels 0; Gaps 0;  
QY 249 AAAGGTTTTCACACAAATTCGTTCCAGGCGCAGCAACATATTTAGTGCAGAGTTCTT 308

Db 24 AGAGGTTTGCACAAATTCGTTCCAGGCGACCGAGCAACATATTTAGTGCAGAAATCTTT 83  
 QY 309 ACCTTGACAAATGATAAACAGATTTGATCTGCTCGGAGTGAAGTAAATATTTATAC 368  
 Db 84 ACCTTGACAAATGATAAACAGATTTGATCTGCTCGGAGTGAAGTAAATATTTATAC 143  
 QY 369 CAACGTTGAGCAAGATGAGAAACCAACAGACAAATGCAATTTACGTTGTCATTTGGAAC 428  
 Db 144 CAACGTTGAGCAAGATGAGAAACCAACAGACAAATGCAATTTACGTTGTCATTTGGAAC 203  
 QY 429 TACTTATGAGATTAAGTACTGACCAATGACCCCTTACACTTTTCTCTCTCTGTTGTTGAAGA 488  
 Db 204 TACTTATGAGATTAAGTACTGACCAATGACCCCTTACACTTTTCTCTCTCTGTTGTTGAAGA 263  
 QY 489 TGGAACTGTTAGTGGTTGATACAGCATCAAACTAGCTGCACAAAGAAAGATTGTAA 548  
 Db 264 TGGAACTGTTAGTGGTTGATACAGCATCAAACTAGCTGCACAAAGAAAGATTGTAA 323  
 QY 549 AGATGATATTTTAACTGTCGACGTCGTCGCCACGCTCTGTTGCTATTTGCCACCAAT 608  
 Db 324 AGATGATATTTTAACTGTCGACGTCGTCGCCACGCTCTGTTGCTATTTGCCACCAAT 383  
 QY 609 ACCATATTTACCTTCTGTTGTTCTGACAGCTCAGTACGAAATATATGATCGGCGAAT 668  
 Db 384 ACCATATTTACCTTCTGTTGTTCTGACAGCTCAGTACGAAATATATGATCGGCGAAT 443  
 QY 669 GCTGGCCACAAGACTACAGGGAATATGACAGTTCGAGGCTACTGGAATGGTTGCCCG 728  
 Db 444 GCTGGCCACAAGACTACAGGGAATATGACAGTTCGAGGCTACTGGAATGGTTGCCCG 503  
 QY 729 TTTTATTCCTTCCCATCTTAATAAATGATCTGACAGTGCATCTCTGTTACAGTGA 788  
 Db 504 TTTTATTCCTTCCCATCTTAATAAATGATCTGACAGTGCATCTCTGTTACAGTGA 563  
 QY 789 AGATGCTCAAGAGATTCGTTAGTACTCTTCAGATACATATCTTTTGACCCGAA 848  
 Db 564 AGATGCTCAAGAGATTCGTTAGTACTCTTCAGATACATATCTTTTGACCCGAA 623  
 QY 849 AGATGATACAGCAGAGAACTTTAAACTCTCTCGGAGAGAGAGAGAGATTGGG 908  
 Db 624 AGATGATACAGCAGAGAACTTTAAACTCTCTCGGAGAGAGAGAGAGAGATTGGG 683  
 QY 909 ACAACCCAGTTAGCGTTGACAGCTTCGTTGATTTGGTCAGATA 955  
 Db 684 GAAAAGCAAGTAAGCGTGCAGACTTCGTTGATTTGGTCAGATA 730

RESULT 18  
 LOCUS BG682657 729 bp mRNA linear EST 01-MAY-2001  
 DEFINITION 602626942F1 NCI\_CGAP\_Skn4 Homo sapiens cDNA clone IMAGE:4751578 5',  
 mRNA sequence.

ACCESSION BG682657  
 VERSION BG682657.1 GI:13914054  
 KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 729)

NIH-MGC <http://mgc.ncl.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

Tissue Procurement: James Cleaver, M.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM10608 row: d column: 11

FEATURES  
 SOURCE

High quality sequence start: 3  
 High quality sequence stop: 693.

Location/Qualifiers  
 1..729

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4751578"

/clone\_lib="NCI\_CGAP\_Skn4"

/tissue\_type="squamous cell carcinoma"

/lab\_host="DH10B (TI phage-resistant)"

/note="Organ: skin; Vector: pCMV-SPORT6; Site: 1: NotI;

Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 1.5kb. Library constructed by Life

Technologies. Note: this is a NCI\_CGAP Library."

BASE COUNT 253 a 165 c 168 g 143 t

ORIGIN

Query Match 26.3%; Score 678.8; DB 12; Length 729;  
 Best Local Similarity 99.3%; Pred. No. 6.2e-181;

Matches 713; Conservative 0; Mismatches 2; Indels 3; Gaps 3;

QY 1394 AGCTGAGCCCCAACCCAGGACAGGTGAACCAAGTTTAAAGTTTGCACTACAGCAGAG 1453

Db 12 ATCTGAGCCCCAACCA -GGACAGGTGAACCAAGTTTAAAGTTTGCACTACAGCAGAG 70

QY 1454 GAACAACCTACAGCACATTAACCTGAACCTTACAGATGAATGGAGCAGTATAGCATCAA 1513

Db 71 GAACAACCTACAGCACATTAACCTGAACCTTACAGATGAATGGAGCAGTATAGCATCAA 130

QY 1514 GTTCTAGAGGAATTTGGAGCCATTGCAATCTGAGGGTCAGGAGGAATCTTTCGTCCAC 1573

Db 131 GTTCTAGAGGAATTTGGAGCCATTGCAATCTGAGGGTCAGGAGGAATCTTTCGTCCAC 190

QY 1574 AGAGTCAGTCAACCCAGCAGAGAGACA -GTGAACAAAAGCTCCTCAAGAAATCA 1632

Db 191 AGAGTCAGTCAACCCAGCAGAGAGAGACA -GTGAACAAAAGCTCCTCAAGAAATCA 250

QY 1633 GAGGATGTGACAAAATATCAGGAAGAGTATCTGCAGAAAACCCAGTTGAGAACCATATC 1692

Db 251 GAGGATGTGACAAAATATCAGGAAGAGTATCTGCAGAAAACCCAGTTGAGAACCATATC 310

QY 1693 AATATAACAATCAGATAAGTTTACAGCCAAAGCCATTGGATTCCAACTCAGGAGAAAGA 1752

Db 311 AATATAACAATCAGATAAGTTTACAGCCAAAGCCATTGGATTCCAACTCAGGAGAAAGA 370

QY 1753 AATGACCTCAATCTTGATCGCTCTTGTGGGTTCCAGAGAAATCTGCTTCATCTGAAAA 1812

Db 371 AATGACCTCAATCTTGATCGCTCTTGTGGGTTCCAGAGAAATCTGCTTCATCTGAAAA 430

QY 1813 GCCAAGAACCAAGAACTTTCAGATCAGACTAGCAGTGCAGAGTGCCTACCAATGAAAAATAC 1872

Db 431 GCCAAGAACCAAGAACTTTCAGATCAGACTAGCAGTGCAGAGTGCCTACCAATGAAAAATAC 490

QY 1873 ACCAATCTGAGCCTCAGTTCNAACAGAGCCACTGGGCTTCAGCTCATGAGAAACA 1932

Db 491 ACCAATCTGAGCCTCAGTTCNAACAGAGCCACTGGGCTTCAGCTCATGAGAAACA 550

QY 1933 TCCACAGGAGCTCTGCTCTTCCAGGACACAGATGACAGTGCATGATGACCCAGTCTGATC 1992

Db 551 TCCACAGGAGCTCTGCTCTTCCAGGACACAGATGACAGTGCATGATGACCCAGTCTGATC 610

QY 1993 CCAGGTGCAAGGTATCGACAGGACCTGGTGTATGAGAGCTCTGCTGTGGCCCTATTTCAG 2052

Db 611 CCAGGTGCAAGGTATCGACAGGACCTGGTGTATGAGAGCTCTGCTGTGGCCCTATTTCAG 670

QY 2053 GAGTTCTTCAGCGGAGAAAGAAAGAAAGAAATGCG -AAGAATTTGGATCTTTGAGC 2109

Db 671 GAGTTCTTCAGCGGAGAAAGAAAGAAAGAAATGCG -AAGAATTTGGATCTTTGAGC 728

RESULT 19

BI464045

LOCUS

783 bp mRNA linear EST 21-AUG-2001

```

DEFINITION 603202819F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5268484 5',
mRNA sequence.
ACCESSION B1464045
VERSION B1464045.1 GI:15254701
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10753 row: f column: 05
High quality sequence stop: 715.
FEATURES             Location/Qualifiers
     source           1..783
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5268484"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Organ: testis; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTNN-3',
size-selected for average insert size 2.2 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT          216 a 170 c 185 g 212 t
ORIGIN
Query Match          26.3%; Score 677.4; DB 13; Length 783;
Best Local Similarity 98.0%; Pred. No. 1.6e-180;
Matches 718; Conservative 0; Mismatches 11; Indels 4; Gaps 3;
QY 1 ATGCTCGGGTGGCTCTACCCACACCTGTTGTGGACGTGAGGAAAAGTCCCTCGGG 60
|||||
DB 41 ATGCTCGGGTGGCTCTACCCACACCTGTTGTGGACGTGAGGAAAAGTCCCTCGGG 100
QY 61 CTGGAGACCCGTCGCCGCTGCGAGTCTGCTACTCTGGGAGAGAGAAATTTATCAAGA 120
|||||
DB 101 CTGGAGACCCGTCGCCGCTGCGAGTCTGCTACTCTGGGAGAGAGAAATTTATCAAGA 160
QY 121 TTAAGACTTGAAGCAACCCCTTAATGTCATGATGTTGTGTTAATCAATCTGTGAAT 180
|||||
DB 161 TTAAGACTTGAAGCAACCCCTTAATGTCATGATGTTGTGTTAATCAATCTGTGAAT 220
QY 181 GACACTGGAGATATATTTATCTGCTCAGATGACACCAAAATAGTAATTAGTAATCCT 240
|||||
DB 221 GACACTGGAGATATATTTATCTGCTCAGATGACACCAAAATAGTAATTAGTAATCCT 280
QY 241 TACAGCAGAAAGTTTGGACAAATTCCTTCAGGCCACCGCAACATATTTAGTGCA 300
|||||
DB 281 TACAGCAGAAAGTTTGGACAAATTCCTTCAGGCCACCGCAACATATTTAGTGCA 340
QY 301 AAGTTCTTACCTTGTACAAATGATAACAGATGTCTCTGCTCTGAGATGAGTAATA 360
|||||
DB 341 AAGTTCTTACCTTGTACAAATGATAACAGATGTCTCTGCTCTGAGATGAGTAATA 400
QY 361 TTTTATACCAAGCTTGAGCAAGATGCAAAACCAACAGACAAATGACGTGTCAT 420

```

```

|||||
Db 401 TTTTATACCAAGCTTGAGCAAGATGCAAAACCAACAGACAAATGACGTGTCAT 460
QY 421 TATGAACCTACTTATGAGATTATGACTGTACCCTTACCCCTTACACATTTCTCTCTGT 480
|||||
Db 461 TATGAACCTACTTATGAGATTATGACTGTACCCTTACCCCTTACACATTTCTCTCTGT 520
QY 481 GGTGAAGATGGAACCTGTAGGTGGTTTGTATACCGCATCAAACTAGCTGCACAAAGAA 540
Db 521 GGTGAAGATGGAACCTGTAGGTGGTTTGTATACCGCATCAAACTAGCTGCACAAAGAA 580
QY 541 GATTCTTA-AAGATGATATTTTAACTGTGACGTGCTGCACGTCTGCTGCTGCTATTTG 599
|||||
Db 581 GATTCTACCAGATGATATTTTAACTGTGACGTGCTGCACGTCTGCTGCTATTTG 640
QY 600 CCCACCATACCATATTTACCTTGTGTTGTTCTGACAGCTCAGTACGAATATATGA 659
Db 641 CCCACCATACCATATTTACCTTGTGTTGTTGTTG-TCTGACAGCTCAGTACGAATATGA 699
QY 660 TCGGCGAATGCTGGG--CACAGAGCTACAGGGAATATGCGAGTTCGAGGACTACTGGA 717
|||||
Db 700 TCGGCGAATGCTGGGCGACAGGAGCTACAGGGAATATGCGGCGTGCAGGCGCTACTGGA 759
QY 718 ATGGTTGCCCGTT 730
|||||
Db 760 CTGGTTGCCCGTT 772

RESULT 20
BG718960 602699258F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4831071 5',
LOCUS mRNA sequence.
DEFINITION BG718960 754 bp mRNA linear EST 08-MAY-2001
ACCESSION BG718960
VERSION BG718960.1 GI:13998147
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10753 row: d column: 16
High quality sequence stop: 749.
FEATURES             Location/Qualifiers
     source           1..754
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4831071"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Organ: testis; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTNN-3',
size-selected for average insert size 2.2 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT          223 a 149 c 167 g 215 t

```

## ORIGIN

Query Match 25.9%; Score 667.8; DB 12; Length 754;  
 Best Local Similarity 95.5%; Pred. No. 8.3e-178;  
 Matches 698; Conservative 0; Mismatches 32; Indels 1; Gaps 1;

QY 210 AGATGACACCAATAGTAATTAATAGTATCTTACAGCAGAGAAGGTTTGCACAAATTCG 269  
 DB 23 AGAGCACTGAGAATACCTTTTAAAGGAGAGGTTCTCATCAGAGTTTGACAAATTCG 82  
 QY 270 TTCAGGSCACCGAGCAACATATTTAGTGCAGAGTTCTTACCTTGTGACAAATGATAACA 329  
 DB 83 TTCAGGSCACCGAGCAACATATTTAGTGCAGAGTTCTTACCTTGTGACAAATGATAACA 142  
 QY 330 GATGTATCTGCTCTGAGAGTGGAGTAATATTTATACCAACGTTGAGCAAGATGAGA 389  
 DB 143 GACTGTATCTGCTCTGAGATGGAGTAATATTTATACCAACGTTGAGCAAGATGAGA 202  
 QY 390 AACCAACAGACAATGCCAATTTAGCTGTCAATATGGAACACTTATGAGATTTAGCTGT 449  
 DB 203 AACCAACAGACAATGCCAATTTAGCTGTCAATATGGAACACTTATGAGATTTAGCTGT 262  
 QY 450 ACCCAATGACCTTACACTTTTCTCTTGTGGAAGATGGAACACTTGTAGGTGGTTGA 509  
 DB 263 ACCCAATGACCTTACACTTTTCTCTTGTGGAAGATGGAACACTTGTAGGTGGTTGA 322  
 QY 510 TACAGCATCAAACTAGCTGCACAAAGAGATTTAAAGATGATATTTAAATTAACG 569  
 DB 323 TACAGCATCAAACTAGCTGCACAAAGAGATTTAAAGATGATATTTAAATTAACG 382  
 QY 570 TCGAGCTGCTGCCAGCTGTGCTATTTGCCCCCAATACCAATATACCTTGTGTTGG 629  
 DB 383 TCGAGCTGCTGCCAGCTGTGCTATTTGCCCCCAATACCAATATACCTTGTGTTGG 442  
 QY 630 TTGTTGACAGCTCAGTACGAATATATGTCGGGATGCTGGCACAAGACTACAGG 689  
 DB 443 TTGTTGACAGCTCAGTACGAATATATGTCGGGATGCTGGCACAAGACTACAGG 502  
 QY 690 GAATATGACAGTGCAGGAGTACTGGAATGTTGCCCGTTTTTATCTTCCCTCTTAA 749  
 DB 503 GAATATGACAGTGCAGGAGTACTGGAATGTTGCCCGTTTTTATCTTCCCTCTTAA 562  
 QY 750 TAATAAGTCTGACAGTGCAGTCTGTTGACAGTGAAGATGTTCAAGAGATCTCGT 809  
 DB 563 TAATAAGTCTGACAGTGCAGTCTGTTGACAGTGAAGATGTTCAAGAGATCTCGT 622  
 QY 810 TAGTTACTCTTCAGATTCATATATCTTTTACCCGAAAGATGATACAGCAGAGAACT 869  
 DB 623 TAGTTACTCTTCAGATTCATATATCTTTTACCCGAAAGATGATACAGCAGAGAACT 682  
 QY 870 TAAACTCTTCTGCGGAGAGAGAGAGAGTTGCGACAAACCCAGTTAAGCGTTT 929  
 DB 683 TAAACTCTTCTGCGGAGAGAGAGAGTTGCGACAAACCCAGTTAAGCGTTT 741  
 QY 930 GAGACTTCGTG 940  
 DB 742 GAGACTTCGTG 752

RESULT 21  
 LOCUS B1461627 834 bp mRNA linear EST 21-AUG-2001  
 DEFINITION 603206744F1 NIH\_MGC\_97 Homo sapiens cDNA clone IMAGE:5272379 5',  
 mRNA sequence.  
 ACCESSION B1461627  
 VERSION B1461627.1 GI:15252283  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 834)  
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE  
JOURNAL  
COMMENT

National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgabs-r@mail.nih.gov](mailto:cgabs-r@mail.nih.gov)  
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
 CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki  
 Toshiyuki and Piero Carninci (RIKEN)  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 Clone Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
 Plate: LLML1687 row: h column: 12  
 High quality sequence start: 5  
 High quality sequence stop: 725.

FEATURES  
source

Location/Qualifiers  
 1..834  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5272379"  
 /clone\_lib="NIH\_MGC\_97"  
 /lab\_host="DH10B"  
 /note="Organ: testis; Vector: pBluescriptR (modified  
 pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI (gtcggg  
 ); Oligo-dr primed using primer 5'-TTTTTTTTTTTTTTVN-3',  
 size-selected for average insert size 2.2 kb and  
 normalized to ROT 5. This is a primary library enriched  
 for full-length clones and constructed using the  
 Cap-trapper method (Carninci, in preparation). Library  
 constructed by M. Brownstein (NIH/NHGRI, National  
 Institutes of Health). Note: this is a NIH\_MGC Library."  
 244 a 163 c 193 g 234 t

BASE COUNT  
ORIGIN

Query Match 25.8%; Score 665.8; DB 13; Length 834;  
 Best Local Similarity 95.2%; Pred. No. 3.2e-177;  
 Matches 719; Conservative 0; Mismatches 32; Indels 4; Gaps 3;

QY 249 AAAGTTTTGACACAATTCGTTTCAGGGCAGCGGACCAACATATTTAGTGCAGAGTTCTT 308  
 DB 24 AGAGTTTTGACACAATTCGTTTCAGGGCAGCGGACCAACATATTTAGTGCAGAGTTCTT 83  
 QY 309 ACCTTGTAACAATGATAACAGATTTGATCTGCTCTGGAGTGGAGTAATATTTATATAC 368  
 DB 84 ACCTTGTAACAATGATAACAGATTTGATCTGCTCTGGAGTGGAGTAATATTTATATAC 143  
 QY 369 CAACGTTGAGCAAGATGCAGAAACCAACAGACAATGCCAATTTAGCTGTCAATATGGAAC 428  
 DB 144 CAACGTTGAGCAAGATGCAGAAACCAACAGACAATGCCAATTTAGCTGTCAATATGGAAC 203  
 QY 429 TACTTATGAGATTTAGCTGTACCAATGACCTTACACTTTCTCTCTGTTGGTGAAGA 488  
 DB 204 TACTTATGAGATTTAGCTGTACCAATGACCTTACACTTTCTCTCTGTTGGTGAAGA 263  
 QY 489 TGGAACTGTTAGTGGTTGTATACAGCATCAAACTAGCTGCACAAAAGAGATTGTA 548  
 DB 264 TGGAACTGTTAGTGGTTGTATACAGCATCAAACTAGCTGCACAAAAGAGATTGTA 323  
 QY 549 AGATGATATTTTAATTAAGTGTGCTGCCAGCTGTCTGCTGTGTTGTTGCCACCAAT 608  
 DB 324 AGATGATATTTTAATTAAGTGTGCTGCCAGCTGTCTGCTGTGTTGTTGCCACCAAT 383  
 QY 609 ACCATATTAACCTGCTGTGTTGTTCTTGACAGCTCAGTACGAATATATGATCGGGAAT 668  
 DB 384 ACCATATTAACCTGCTGTGTTGTTCTTGACAGCTCAGTACGAATATATGATCGGGAAT 443  
 QY 669 GCTGGGCAAGAGCTACAGGAATATGACAGTCCAGGAGTACTTGGAAATGTTGCCCG 728  
 DB 444 GCTGGGCAAGAGCTACAGGAATATGACAGTCCAGGAGTACTTGGAAATGTTGCCCG 503  
 QY 729 TTTTATCTTCCCATCTTTAATAAAGTCTTCGAGAGTGCATCTCTGTTACAGTGA 788  
 DB 504 TTTTATCTTCCCATCTTTAATAAAGTCTTCGAGAGTGCATCTCTGTTACAGTGA 563

```

QY 789 AGATGGTCAAGAGATCTCGTTAGTTACTCTTCAGATTACATATATCTTTTGGACCCGAA 848
DB 564 AGATGGTCAAGAGATCTCGTTAGTTACTCTTCAGATTACATATATCTTTTGGACCCGAA 623
QY 849 AGATGATACAGACAGAGACTTAAACTCTCTCTGCGGAAGAG-AGAAGAGAGAGTTGC 907
DB 624 AGATGATCCGACGAGAACTTAAACTCTCTCTGCGGAAGAGAGAGAGAGTTGC 683
QY 908 G-ACAACCACCAGTTTAAAGGTTTGAAGTTCGTTGTA--TTGGTCAGATACCTGGACCCA 964
DB 684 GAACAACCACCAGTTTAAAGGTTTGAAGTTCGTTGTA--TTGGTCAGATACCTGGACCCA 743
QY 965 GAGCAAGCCGAGAGTGAACGAGACGAGATGGA 999
DB 744 GAAGCATGCCGGAAGTACCAGAGACGAGAAGGA 778

RESULT 22
LOCUS BI464622 850 bp mRNA linear EST 21-AUG-2001
DEFINITION 603203236f1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5268920 5',
mRNA sequence.
ACCESSION BI464622
VERSION BI464622.1 GI:15255278
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 850)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: L14M11678 row: h column: 09
High quality sequence stop: 710.
FEATURES
source
location/Qualifiers
1..850
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5268920"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/note="Organ: testis; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTIVN-3',
size-selected for average insert size 2.2 kb and
normalized to 80% 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NHGRI/NIH, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 243 a 182 c 188 g 237 t
ORIGIN
Query Match 25.7%; Score 663.6; DB 13; Length 850;
Best Local Similarity 93.0%; Pred. No. 1.4e-176;
Matches 785; Conservative 0; Mismatches 44; Indels 15; Gaps 8;
QY 1 ATCTCTCGGGTGGCTCTACCCACACCTGTTGGGACGTGAGAAAGTCCCTCGG 60
DB 6 ATCTCTCGGGTGGCTCTACCCACACCTGTTGGGACGTGAGAAAGTCCCTCGG 65

```

```

61 CTGAGGACCCGTCCTCCGCTCGGAGTCGCTACCTGGGGAAGAGAAATTTATCCAAAGA 120
66 CTGAGGACCCGTCCTCCGCTCGGAGTCGCTACCTGGGGAAGAGAAATTTATCCAAAGA 125
121 TTAATACTTGAAGCAACCCCTTAATGTGCATGATGGTGTGTTAATAACAATCTGTGGAAT 180
126 TTAATACTTGAAGCAACCCCTTAATGTGCATGATGGTGTGTTAATAACAATCTGTGGAAT 185
181 GACACTGGAGATATATTTATCTGGCTCAGATGACACCAAAATAGTAATAGTAATCTCT 240
186 GACACTGGAGATATATTTATCTGGCTCAGATGACACCAAAATAGTAATAGTAATCTCT 245
241 TACAGCAAAAGGTTTTCACAACAATTCGTTCCAGGSCACCGAGCAACATATTAGTGCA 300
246 TACAGCAAAAGGTTTTCACAACAATTCGTTCCAGGSCACCGAGCAACATATTAGTGCA 305
301 AAGTCTTACCTTGTACAAATGATAAAGATGTTATCTCTCTGGAGATGAGTAATA 360
306 AAGTCTTACCTTGTACAAATGATAAAGATGTTATCTCTCTGGAGATGAGTAATA 365
361 TTTTATACCAACGTTGAGCAAGATGCAGAAACCAACACAGCAATGCCAATTTACGTGCA 420
366 TTTTATACCAACGTTGAGCAAGATGCAGAAACCAACAGCAATGCCAATTTACGTGCA 425
421 TATGGAACCTACTTATGAGATTATGACTGTACCCATGACCCCTTACACTTTTCTCTTCT 480
426 TATGGAACCTACTTATGAGATTATGACTGTACCCATGACCCCTTACACTTTTCTCTTCT 485
481 GGTGAAGATGGAATGTTAGTGGTTGATACACGATCAAAACTAGCTGCACAA--A 536
486 GGTGAAGATGGAATGTTAGTGGTTGATACACGATCAAAACTAGCTGCACAAACGA 544
537 AGAAGATTGTAAGACATATTTAATAA--TGTCCAGCTGCTGCACAGCTGTGCT 594
545 AGATTGTACACAGATGCTTTTAAATAAAGCTGCTGCACAGCTGTGCT 604
595 ATTGCCCACCAATACATATTTACCTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 654
605 ATTGCCCACCAATACATATTTACCTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 664
655 TATGTCGGGGAATGCTGGGAC--AAGAGCTACAGGG--AATTATGACGG--TCGAGGGAC 710
665 TATGTCGGGGAATGCTGGGAC--AAGAGCTACAGGG--AATTATGACGG--TCGAGGGAC 724
711 -TACTGGAATGGTGGCCGTTTATTCCTCCATCTTAATAAAGCTTCGACAGTGA 769
725 ATACTGGAATGGTGGCCGTTTATTCCTCCATCTTAATAAAGCTTCGACAGTGA 784
770 CATCTCTGTTTACA---GTGAAGATGGTCAAGAGATTCCTGTTAGTTACTCTTCAGATT 826
785 ACTCTCTGTTTACACACTGCAGATGGTCAACAAATCCTCGTAATTCACCTCCCATAT 844
827 ACAT 830
845 ACTT 848

```

```

RESULT 23
LOCUS BE180558 694 bp mRNA linear EST 22-JUN-2000
DEFINITION RC3-HT0625-130400-021-c02 HT0625 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE180558
VERSION BE180558.1 GI:8659734
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 694)
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovsky-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare

```



M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.J.  
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.J.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-RC3-HT0625-130400-021-c02&st3-2000-04-13&st4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 678.

```
FEATURES             source
  1..694
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone_lib="HT0625"
    /dev_stage="Adult"
    /note="Organ: head neck; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORFES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT          208 a 136 c 143 g 207 t
ORIGIN
Query Match          25.5%; Score 558.2; DB 10; Length 694;
Best Local Similarity 99.3%; Pred. No. 4.2e-175;
Matches 682; Conservative 0; Mismatches 3; Indels 2; Gaps 2;
QY 187 GGAGATATATTTATCTGCTGATGACACACCAATAGTAATAGTAATCCCTTACAGC 246
DB 1 CGAGATATATTTATCTGCTGATGACACCAATAGTAATAGTAATCCCTTACAGC 60
QY 247 AGAAGGTTTGGACAAATTCGTTGAGGGCAGCAGCAACATATTTAGTGCAAGTTC 306
DB 61 AGAAGG-TTTGACAAATTCGTTGAGGGCAGCAGCAACATATTTAGTGCAAGTTC 119
QY 307 TTACCTTGATGATATGATACAGATTTGATCTGCTGCTGAGATGGAGTATATTTAT 366
DB 120 TTACCTTGATGATATGATACAGATTTGATCTGCTGCTGAGATGGAGTATATTTAT 179
QY 367 ACCAAGTTGAGCAAGATGCAAGAACCAACACAGCAATGCCAATTTACGTGTCATTATGA 426
DB 180 ACCAAGTTGAGCAAGATGCAAGAACCAACACAGCAATGCCAATTTACGTGTCATTATGA 239
QY 427 ACTACTTATGAGATTTAGCTGTACCAATGACCTTACACTTTTCTCTCTCTGCTGCTGGA 486
DB 240 ACTACTTATGAGATTTAGCTGTACCAATGACCTTACACTTTTCTCTCTCTGCTGCTGGA 299
QY 487 GATGCACTGTTAGTGTTGATACAGCAATCAAACTAGCTGCAACAAAGAGATTTGT 546
DB 300 GATGCACTGTTAGTGTTGATACAGCAATCAAACTAGCTGCAACAAAGAGATTTGT 359
QY 547 AAAGATGATATTTTAAATTAACGTGCGAGCTGCTGCCACGCTGTTGCTATTTGCCACCA 606
DB 360 AAAGATGATATTTTAAATTAACGTGCGAGCTGCTGCCACGCTGTTGCTATTTGCCACCA 419
QY 607 ATACCATATTTACTTCTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 666
DB 420 ATACCATATTTACTTCTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 479
QY 667 ATGCTGGGCACAAAGAGCTACAGGAATTTATGAGGCTGAGGGGACTACTGGAATGGTTGCC 726
```

```
Db 480 ATGCTGGGCACAAAGAGCTACAGGAATTTATGAGGCTGAGGGACTACTGGAATGGTTGCC 539
QY 727 CGTTTATTCCTTCCCATCTTAATAAGTCCGAGAGTGACATCTCTGTTGTACAGT 786
DB 540 CGTTTATTCCTTCCCATCTTAATAAGTCCGAGAGTGACATCTCTGTTGTACAGT 599
QY 787 GAGATGCTCAAGAGATTCCTGTTAGTTACTTTCAGATTACATATATCTTTTGGACCG 846
DB 600 GAGATGCTCAAGAGATTCCTGTTAGTTACTTTCAGATTACATATATCTTTTGGACCG 659
QY 847 AAGATGATACAGCAGCAGAGACTTAAA 873
DB 660 AAGATGAT-CAGCAGCAGAGACTTAAA 685

RESULT 24
BG117357 602346361F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4441431 5',
LOCUS mRNA sequence.
DEFINITION BG117357
ACCESSION BG117357.1 GI:12610863
VERSION BG117357
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10212 row: 1 column: 16
High quality sequence stop: 657.

FEATURES             source
  1..739
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone_lib="IMAGE:4441431"
    /clone_lib="NIH_MGC_90"
    /tissue_type="adenocarcinoma, cell line"
    /lab_host="DH10B (phage-resistant)"
    /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.7 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."
BASE COUNT          222 a 143 c 177 g 197 t
ORIGIN
Query Match          25.4%; Score 655.8; DB 12; Length 739;
Best Local Similarity 97.7%; Pred. No. 2.1e-174;
Matches 708; Conservative 0; Mismatches 12; Indels 5; Gaps 4;
QY 538 GAAGATTGTAAGATCATATTTTAACTCTGACAGTGCTGCCAGCTCTGTTGCTATT 597
DB 1 GAAGATTGTAAGATCATATTTTAACTCTGACAGTGCTGCCAGCTCTGTTGCTATT 60
QY 598 TGCCCAACCAATACCATATTTACCTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 657
DB 61 TGCCCAACCAATACCATATTTACCTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 120
QY 658 GATCGCGCAATGCTGGGCAAGAGCTACAGGAATTTATGAGGCTGAGGGGACTACTGGA 717
```



```

Db 121 GATCGCGCAATCTGGGCAAGAGCTACAGGAATTAATGAGGTCGAGGGACTACTGGA 180
QY 718 ATGGTGGCCGTTTATCTCCATCTTAATAATAAGTCTCTGAGAGTGAGCATCTCTG 777
Db 181 ATGGTGGCCGTTTATCTCCATCTTAATAATAAGTCTCTGAGAGTGAGCATCTCTG 240
QY 778 TCTTACAGTGAAGTGGTCAAGAGATTCCTGTTAGTACTCTTCAGATTACATATCTT 837
Db 241 TCTTACAGTGAAGTGGTCAAGAGATTCCTGTTAGTACTCTTCAGATTACATATCTT 300
QY 838 TTGACCCGAAAGATGATACACGAGAGAACTTAAACCTCTCTGCGGAAGAGAGA 897
Db 301 TTGACCCGAAAGATGATACACGAGAGAACTTAAACCTCTCTGCGGAAGAGAGA 360
QY 898 GAAGAGTGGCGACACACAGCTTAAGCGTTTGAGACTTCGTTGGTGAATGGTCA 957
Db 361 GAAGAGTGGCGACACACAGCTTAAGCGTTTGAGACTTCGTTGGTGAATGGTCA 420
QY 958 GGACCCAGAGCAAGCGGAGAGTGAACGAGAGAGATGGAGACAGATCCCAATGTG 1017
Db 421 GGACCCAGAGCAAGCGGAGAGTGAACGAGAGAGATGGAGACAGATCCCAATGTG 480
QY 1018 TCATTGATGACAGAAATCTCTGATATGTTATCAAGATGTTTGAAGAAGAGTGAGTT 1077
Db 481 TCATTGATGACAGAAATCTCTGATATGTTATCAAGATGTTTGAAGAAGAGTGAGTT 540
QY 1078 GCACAAAGCAATAGAGGACGAGGAGATCTCGACCCAGAGTGGAAACAAATCAGAT 1137
Db 541 GCACAAAGCAATAGAGGACGAGGAGATCTCGACCCAGAGTGGAAACAAATCAGAT 600
QY 1138 ATTTCAACTCTTCTACCGTCCCATCAAGTCTCTGATTTGGAAGTGAAGTCAATG 1197
Db 601 ATTTCAACTCTTCTACCGTCCCATCAAGTCTCTGATTTGGAAGTGAAGTCAATG 658
QY 1198 GAAGTAGATCTCAGCTGAACAAATTTCTTCAGCCTTCTACATCTCTACAAATGTCAGCT 1257
Db 659 GACGTAAATACCTCCAG-TGAACAATTTCTTCAGC--TTCTAATCTCTCAAGGTCAGTC 715
QY 1258 CAGGC 1262
Db 716 AAGGC 720

RESULT 25
BM923018
LOCUS AGENCOURT_6632140 NIH_MGC_118 Homo sapiens cDNA clone IMAGE:5756969
DEFINITION 5', mRNA sequence.
ACCESSION BM923018
VERSION BM923018.1 GI:19373397
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE NTH-MGC http://mgc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@p5r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LNLN)
Clone Distribution: Agencourt Bioscience Corporation
found through the I.M.A.G.E. Consortium information can be
http://image.lnl.gov
plate: L1AM12798 row: g column: 18
High quality sequence start: 14
High quality sequence stop: 693.
Location/Qualifiers
1. .1075

FEATURES
source

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5756969"
/clone_lib="NIH_MGC_118"
/tissue_type="leukocyte"
/lab_host="DH108"
/notes="Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV
(deprecated); RNA source leukocytes from anonymous pool of
non-activated adult donors. Library is oligo-dT primed
and directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.7 kb, insert size range
1.2-3.3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(invitrogen). Research Genetics tracking code 027. Note:
this is a NIH_MGC Library."
BASE COUNT 277 a 264 c 267 g 267 t
ORIGIN
Query Match 25.1%; Score 648.6; DB 14; Length 1075;
Best Local Similarity 95.8%; Pred. No. 2.8e-172;
Matches 731; Conservative 0; Mismatches 24; Indels 8; Gaps 6;
QY 1 ATGTCTCGGGTGGCTCTTACCCACACCTGTTGTGGAGCGTGAGGAAAGTCCCTCGG 60
Db 226 ATGTCTCGGGTGGCTCTTACCCACACCTGTTGTGGAGCGTGAGGAAAGTCCCTCGG 285
QY 61 CTGGAGGACCGCTCCCGGCTGGAGTCGCTACCTCGGAGAGAGAGAAATTTATCCAA 120
Db 286 CTGGAGGACCGCTCCCGGCTGGAGTCGCTACCTCGGAGAGAGAGAAATTTATCCAA 345
QY 121 TTAACACTTGAAGCAACCTTAATGTCATGATGTTGTTTAAATACAACTCTGTTGGA 180
Db 346 TTAACACTTGAAGCAACCTTAATGTCATGATGTTGTTTAAATACAACTCTGTTGGA 405
QY 181 GACACTGAGGATATATTTATCTGGCTCAGATGACACCAATTTAGTAAATCT 240
Db 406 GACACTGAGGATATATTTATCTGGCTCAGATGACACCAATTTAGTAAATCT 465
QY 241 TACAGCAAGAGGTTTGAACAACAAATTCGTTCCAGGCGCCGAGCAACATATTAGTGA 300
Db 466 TACAGCAAGAGGTTTGAACAACAAATTCGTTCCAGGCGCCGAGCAACATATTAGTGA 525
QY 301 AAGTTCTTACCTTGPACAAATGATAACAGATTTGATCTCTCTGGAGATGAGTAATA 360
Db 526 AAGTTCTTACCTTGPACAAATGATAACAGATTTGATCTCTCTGGAGATGAGTAATA 585
QY 361 TTTATACCAACGTTGACCAAGATCCAGAAACCAACACAGCAATGCCAATTTACGTG 420
Db 586 TTTATACCAACGTTGACCAAGATCCAGAAACCAACACAGCAATGCCAATTTACGTG 645
QY 421 TATGGAATCTATGAGATTATGACTGTACCCAAATGA-CCCTTACACTTTTCTCTCTG 479
Db 646 TATGGAATCTATGAGATTATGACTGTACCCAAATGACCCCTTACACTTTTCTCTCTG 705
QY 480 TGGTGAAGATGGAATCTTTAGGTTGATGACGCAATCAAACTAGCTGCACAAAAGA 539
Db 706 TGGTGAAGATGGAATCTTTAGGTTGATGACGCAATCAAACTAGCTGCACAAAAGA 765
QY 540 AGATTGTAAGATGATATTTAAATTAAGTGTGACGCTGCTCCAGCTCTGTGTTAT 599
Db 766 AGATTGTAAGATGATATTTAAATTAAGTGTGACGCTGCTCCAGCTCTGTGTTAT 825
QY 600 CCCACCAATACCATATTTACCTTGTCTGTTGTTT-GTTCTGACAGCTCAGTACGATATATG 658
Db 826 CCCACCAATACCATATTTACCTTGTCTGTTGTTGTTGTTCTGACAGCTCCGTAATATATG 885
QY 659 AT-CGGGAATGCTGGGACAAAG-AGCTACAGGAAATTTATCAGGTGCGA--GGGACTACT 714
Db 886 ATCCGGGAATGCTGGGACAAAGAACCTCCCGGGAATTTATCAGGCGCAAGGAAATACCT 945
QY 715 GGAATGCTG--CCCGTTTATCTCTTCCCATCTTATAATAA 755
Db 946 GGAATGCTGCGCGGTTTAAATCCCTTCCCATCTTAAAAATAA 988

```

```

RESULT 26
BE867206
LOCUS
DEFINITION
  BE867206 775 bp mRNA linear EST 20-OCT-2000
  601442266F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3846472 5',
  mRNA sequence.
ACCESSION
  BE867206
VERSION
  BE867206.1 GI:10315982
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 775)
  NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS
  National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE
  Unpublished (1999)
COMMENT
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs-remail.nih.gov
  Tissue Procurement: ATCC
  cDNA Library Preparation: Life Technologies, Inc.
  cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
  DNA Sequencing by: Incyte Genomics, Inc.
  Clone Distribution: MGC clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at:
  http://image.llnl.gov
  Plate: LLAM9559 row: c column: 17
  High quality sequence stop: 611.
  Location/Qualifiers
    1..775
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="IMAGE:3846472"
    /clone_lib="NIH_MGC_65"
    /tissue_type="adenocarcinoma"
    /lab_host="DH10B (phage-resistant)"
    /note="Organ: colon; Vector: pCMV-Sport6; Site_1: NotI;
    Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
    Average insert size 1.8 kb. Library constructed by Life
    Technologies."
BASE COUNT 255 a 182 c 186 g 152 t
ORIGIN
  Query Match 25.1%; Score 646.6; DB 12; Length 775;
  Best Local Similarity 94.4%; Pred. NO. 8.6e-172;
  Matches 694; Conservative 0; Mismatches 34; Indels 7; Gaps 2;
QY 909 ACAACCCAGTTAAGCGTTTCAGACTTCGTGGTGGTTCAGATCTGGTCCAGATCTGGACCCAGAGC 968
Db 1 ACAACCCAGTTAAGCGTTTCAGACTTCGTGGTGGTTCAGATCTGGTCCAGATCTGGACCCAGAGC 60
QY 969 AAGCGCGGAGAGTGAACGAGACGAGATGGAGAGCAGAGTCCCAATGTCTATTGATGCA 1028
Db 61 AAGCGCGGAGAGTGAACGAGACGAGATGGAGAGCAGAGTCCCAATGTCTATTGATGCA 120
QY 1029 GAGAATGCTGATATGTTATCAAGATGGTTTGAAGAGCAAGTGGTTGGCAACAAGCAA 1088
Db 121 GAGAATGCTGATATGTTATCAAGATGGTTTGAAGAGCAAGTGGTTGGCAACAAGCAA 180
QY 1089 TAGAGGACGAGGAGATCTCGACCCAGAGTGGACAAGTCAATCAGATATTTCAACTCT 1148
Db 181 TAGAGGACGAGGAGATCTCGACCCAGAGTGGACAAGTCAATCAGATATTTCAACTCT 240
QY 1149 TCCTACGGTCCCATCAAGTCTCTGATTTGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1208
Db 241 TCCTACGGTCCCATCAAGTCTCTGATTTGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 300
QY 1209 TCAGGTGAACAATTTCTCAGCCTTCTACATCCTCTACAAATGTCAAGTCAAGTCAAGTCA 1268
Db 301 TCAGGTGAACAATTTCTCAGCCTTCTACATCCTCTACAAATGTCAAGTCAAGTCAAGTCA 360
QY 1269 GACATCATCTCCACAGAAAGCCCTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1328

```

```

Db 361 GACATCATCTCCACAGAAAGCCCTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 420
QY 1329 ACAAGGAGCTCTGTGAGGCATCTGGACACACACACATCATCATCTGATGATAAATAA 1388
Db 421 ACAAGGAGCTCTGTGAGGCATCTGGACACACACACATCATCATCTGATGATAAATAA 480
QY 1389 TGAAGAAGCTGAGCCCAACACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1448
Db 481 TGAAGAAGCTGAGCCCAACACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 540
QY 1449 AGAAGGAGCAACTCAAGCAACAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1508
Db 541 AGAAGGAGCAACTCAAGCAACAATAAATAAATAAATAAATAAATAAATAAATAAATAA 600
QY 1509 ATCAAGTTCTAGAGGAAT--GGGAGCCATTGCAATCTGAGGGTCAGGAGG-----AAT 1561
Db 601 ATCAAGTTCTAGAGGAATCGGGAATCGGGAATCGGGAATCGGGAATCGGGAATCGGGA 660
QY 1562 CTTTTCCTCCACAGAGCTCAGTCAACACACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1621
Db 661 CGTCCCAACAGAGCTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 720
QY 1622 AAGATCATCAGAGG 1636
Db 721 GAGAACCCACCAAG 735
RESULT 27
LOCUS
  BG772319 806 bp mRNA linear EST 15-MAY-2001
DEFINITION
  602722419F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4839173 5',
  mRNA sequence.
ACCESSION
  BG772319
VERSION
  BG772319.1 GI:14082972
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 806)
  NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS
  National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE
  Unpublished (1999)
JOURNAL
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs-remail.nih.gov
  Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
  cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
  Toshiyuki and Piero Carninci (RIKEN)
  cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
  DNA Sequencing by: Incyte Genomics, Inc.
  Clone Distribution: MGC clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at:
  http://image.llnl.gov
  Plate: LLAM10774 row: f column: 06
  High quality sequence stop: 791.
  Location/Qualifiers
    1..806
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="IMAGE:4839173"
    /clone_lib="NIH_MGC_97"
    /lab_host="DH10B"
    /note="Organ: testis; Vector: pBluescriptR (modified
    pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcag
    ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTNN-3',
    size-selected for average insert size 2.2 kb and
    normalized to 5. This is a primary library enriched
    for full-length clones and constructed using the
    Cap-trapper method (Carninci, in preparation). Library
    constructed by M. Brownstein (NIH/NHGRI, National
    Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 230 a 169 c 181 g 226 t

```

## ORIGIN

Query Match 24.9%; Score 641.8; DB 12; Length 806;  
 Best Local Similarity 99.1%; Pred. No. 2e-170;  
 Matches 687; Conservative 0; Mismatches 2; Indels 4; Gaps 4;

QY 97 GAAGAAGAGAGATTTATCAAGATTAATAAATTTGAAGCAACCCCTTAATGTCATGATGGT 156  
 Db 105 GGAAGAAGAGATTTATCAAGATTAATAAATTTGAAGCAACCCCTTAATGTCATGATGGT 164

QY 157 TGTGTTAATACATCTGTTGGGAATGACACTGGAGATATATTTATCTGCTCAGATGAC 216  
 Db 165 TGTGTTAATACATCTGTTGGGAATGACACTGGAGATATATTTATCTGCTCAGATGAC 224

QY 217 ACCAATAGTAGTAATAGTAATCCTTACAGCAGAAAGGTTTTGACAAACATTCGTTTCAGGG 276  
 Db 225 ACCAATAGTAGTAATAGTAATCCTTACAGCAGAAAGGTTTTGACAAACATTCGTTTCAGGG 284

QY 277 CACCAGCAACATATTTAGTGCAGAAAGTCTTTACCTTGTACAAATGATAACAGATGTA 336  
 Db 285 CACCAGCAACATATTTAGTGCAGAAAGTCTTTACCTTGTACAAATGATAACAGATGTA 344

QY 337 TCTCTCTGGAGATGAGTATATTTATACCAACGTTGAGCAAGATGCAGAAACCAAC 396  
 Db 345 TCTCTCTGGAGATGAGTATATTTATACCAACGTTGAGCAAGATGCAGAAACCAAC 404

QY 397 AGACAATGCCAATTTACGTGTCATTTAGTGAATGATGATGATGATGATGATGATGAT 456  
 Db 405 AGACAATGCCAATTTACGTGTCATTTAGTGAATGATGATGATGATGATGATGATGAT 464

QY 457 GACCTTTACATCTTCTCTCTGTTGTTGCAAGATGCAAGTGTAGTGTGTTGATACAGC 516  
 Db 465 GACCTTTACATCTTCTCTCTGTTGTTGCAAGATGCAAGTGTAGTGTGTTGATACAGC 524

QY 517 ATCAAACTAGTGCACAAAAGAGATGTAATGATGATATTTAATTAATTAATTAATTAAT 575  
 Db 525 ATCAAACTAGTGCACAAAAGAGATGTAATGATGATATTTAATTAATTAATTAATTAAT 584

QY 576 TGCTGCCACGCTGTTGCTATTTCCCAACCAATACCATATTAACCTTGTGTTGTTGTC 635  
 Db 585 TGCTGCCACGCTGTTGCTATTTCCCAACCAATACCATATTAACCTTGTGTTGTTGTC 644

QY 636 TGACAGCTCAGTACGAATATATGATCGCGCAATGCTGGGCAAGAGCTACAGGGAATTA 695  
 Db 645 TGACAGCTCAGTACGAATATATGATCGCGCAATGCTGGGCAAGAGCTACAGGGAATTA 704

QY 696 TGCAGGTCGAGGACTACTGGAA-TGGTGGCCCGTTTATTCCTTCCC-AFTCTTAATAAT 753  
 Db 705 TGCAGGTCGAGGACTACTGGAA-TGGTGGCCCGTTTATTCCTTCCC-AFTCTTAATAAT 764

QY 754 AAGTCCTG-CAGAGTGACATCTCTGTGTTACAG 785  
 Db 765 AAGTCCTGCCAGAGTGACATCTCTGTGTTACAG 797

## RESULT 28

RG530857  
 LOCUS 602559894F1 NIH\_MGC\_61 Homo sapiens cDNA clone IMAGE:469798 5',  
 mRNA sequence.

DEFINITION RG530857.1 GI:13522394

ACCESSION RG530857

VERSION EST.

KEYWORDS human.

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 911)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

CONTACT Contact: Robert Strausberg, Ph.D.

COMMENT Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

## Tissue Procurement: ATCC

cDNA Library Preparation: CLONETECH Laboratories, Inc.  
 DNA Sequencing Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 Clone Distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>

Plate: LLCM1527 row: k column: 23

High quality sequence stop: 701.

FEATURES  
source

Location/Qualifiers  
 1..911  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:469798"  
 /clone\_lib="NIH\_MGC\_61"  
 /tissue\_type="embryonal carcinoma"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: testis; Vector: pDNR-LIB (Clontech); Site\_1:  
 SfiI (ggccattggcc); Site\_2: SfiI (ggccattggcc);  
 Double-stranded cDNA was prepared from cell line RNA.  
 and 3' adaptors were used in cloning as follows: 5'  
 adaptor sequence: 5'-CAGCGCATATGCGC-3' and 3' adaptor  
 sequence: 5'-ATTCTAGAGCGGCGGCGGAGATG-dt(30)BN-3'  
 (where B = A, C, or G and N = A, C, G, or T). Average  
 insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies  
 contained inserts by PCR. This library was enriched for  
 full-length clones and was constructed by Clontech  
 Laboratories (Palo Alto, CA). Note: this is a NIH\_MGC  
 Library."

BASE COUNT 310 a 168 c 214 g 219 t

Query Match 24.9%; Score 641.6; DB 12; Length 911;  
 Best Local Similarity 90.1%; Pred. No. 2.5e-170;  
 Matches 762; Conservative 0; Mismatches 34; Indels 50; Gaps 5;

QY 1676 CAGTTGAGAACCATATCAATATACACAAATCAGATAAGTTACACGCAAGCCATGGATT 1735  
 Db 1 CAGTTGAGAACCATATCAATATACACAAATCAGATAAGTTACACGCAAGCCATGGATT 60

QY 1736 CCAACTCAGGAGAAAGAAATGACCTCAATCTTGATCGCTCTTGTGGGTTCCAGAAAGAT 1795  
 Db 61 CCAACTCAGGAGAAAGAAATGACCTCAATCTTGATCGCTCTTGTGGGTTCCAGAAAGAT 120

QY 1796 CTGCTTCATCTGAAAAAGCCAAAGAACCAAGAACTTCAGATCAGACTAGCAGTGAAGTG 1855  
 Db 121 CTGCTTCATCTGAAAAAGCCAAAGAACCAAGAACTTCAGATCAGACTAGCAGTGAAGTG 180

QY 1856 CTACCAATGAAAAATACACCAATCCTGAGCCTCAGTTCCAAACAGAGCCACTGGGCCTT 1915  
 Db 181 CTACCAATGAAAAATACACCAATCCTGAGCCTCAGTTCCAAACAGAGCCACTGGGCCTT 240

QY 1916 CAGCTCATGAAGAAACATCCACCGGACTCTGCTCTTCAGGACACAGATGACAGTGAATG 1975  
 Db 241 CAGCTCATGAAGAAACATCCACCGGACTCTGCTCTTCAGGACACAGATGACAGTGAATG 300

QY 1976 ATGACCCAGTCTCTGATCCCAAGGTATCGAGCAGGACCTGGTGTATAGACGCTCTG 2035  
 Db 301 ATGA-----TGATAGACGCTCTG 318

QY 2036 CTGTTGCCCCGATTCAGGAGTCTTCAGACGGAGAAAGAAAGAAATGGAAGAAAT 2095  
 Db 319 CTGTTGCCCCGATTCAGGAGTCTTCAGACGGAGAAAGAAAGAAATGGAAGAAAT 378

QY 2096 TGGATACCTTTGAACATTAGAAGCGCGCTAGTAAATAATGGTTTATAAAGCCATCGCAACT 2155  
 Db 379 TGGATACCTTTGAACATTAGAAGCGCGCTAGTAAATAATGGTTTATAAAGCCATCGCAACT 438

QY 2156 CCAGGCAATGATAAAGAGCCCAATTTCTGGGTGCTAACTTTGTATAGTGGTCTG 2215  
 Db 439 CCAGGCAATGATAAAGAGCCCAATTTCTGGGTGCTAACTTTGTATAGTGGTCTG 498

QY 2216 ACTGTGCCCCACATTTTCATCTGGATCGGCACACTGCTGACATTTGATGCTTCTGGAAG 2275

|||||  
Db 499 ACTGTGCCACATTTTCATCTGGGATCGGCACACTGCTGAGCATTTTGATGCTTGTGAAG 558  
|||||  
QY 2276 CTGATAATCATGTGGTAAATCGCTGAGCCACATCGTTTGACCCCAATTT-TAGCCCTCA 2334  
|||||  
Db 559 CTGATAATCATGTGGTAAATCGCTGAGCCACATCGTTTGACCCCAATTTTGATGCTTGTGAAG 618  
|||||  
QY 2335 TCTGGCATAGATTATGACATAAAGATCTGTGTCACCA- - -TTAGAAGAGTCAAGGATTTT 2390  
|||||  
Db 619 TCTGGCATAGATTATGACATAAAGATCTGTGTCACCAATATAGAGAGTCAAGGATTTT 678  
|||||  
QY 2391 TAACCGAAAATCTGCTGATGAAGTT-ATAACTCGAAGCAAGCACTCATGCTGGAAGAACTA 2449  
|||||  
Db 679 TAACCGAAGAAATCTGCTGATGAAGTTAATAGTCGAAGCAAGCACTCATGATGGAAGAACTA 738  
|||||  
QY 2450 GAACACCAATACAGTTCCAGCCCTTTCATGTTGAGGATGTTGCTTCACTTAAATCATATA 2509  
|||||  
Db 739 GAACCAAC- -AATAAGTTCCAAAGCATATTTATGTTGAGGATGTGGGTTCACTTAAATCATATA 796  
|||||  
QY 2510 TCCGAG 2515  
|||||  
Db 797 TTCGAG 802  
|||||

RESULT 29  
BQ019842/c  
LOCUS  
DEFINITION BQ019842 648 bp mRNA linear EST 27-MAR-2002  
IMAGE:5827567 3', mRNA sequence.  
ACCESSION BQ019842  
VERSION BQ019842.1 GI:19755119  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 648)  
NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgap@remail.nih.gov](mailto:cgap@remail.nih.gov)  
Tissue Procurement: Dr. Jose Mercuende  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Seq primer: M13 FORWARD  
POLYA=Yes.

Location/Qualifiers  
1. 648  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5827567"  
/tissue\_type="Chondrosarcoma"  
/dev\_stage="Adult"  
/lab\_host="DH10B (Life Technologies)"  
/note="Organ: Left Pubic Bone; Vector: pT7T3-Pac  
(Pharmacia) with a modified polylinker; Site 1: EcoR I;  
Site 2: Not I; NCI-CCGAP-ED0 is a cDNA library containing  
the following tissue(s): Chondrosarcoma cell line C55. The  
library was constructed according to Bonaldo, Lennon and  
Soares, Genome Research, 6:791-806, 1996. First strand  
cDNA synthesis was primed with an oligo-dT primer  
containing a Not I site. Double stranded cDNA was ligated  
to an EcoR I adaptor, digested with Not I, and cloned  
directionally into pT7T3-Pac vector. The oligonucleotide  
used to prime the synthesis of first-strand cDNA contains  
a library tag sequence that is located between the Not I  
site and the (dT)18 tail. The sequence tag for this

Library is GCTCAAGGCT.

TAG\_LIB=UI-H-ED0

TAG\_TISSUE=chondrosarcoma

TAG\_SEQ=GCTCAAGGCT"

BASE COUNT 123 a 143 c 148 g 234 t

ORIGIN

Query Match 24.8%; Score 640; DB 14; Length 648;  
Best Local Similarity 99.2%; Pred. No. 5.9e-170;  
Matches 643; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1438 CACTACAGCAGAGGAAGCAACATACAGACACATAAATGAACTTACAGATGAATGG 1497  
|||||  
Db 648 CACTACAGCAGAGGAAGCAACATACAGACACATAAATGAACTTACAGATGAATGG 589  
|||||  
QY 1498 AGCAGTATAGCATCAAGTTCTAGAGGAATTTGGAGCCCAATTCGAAATCTGAGGGTCAGAG 1557  
|||||  
Db 588 AGCAGTATAGCATCAAGTTCTAGAGGAATTTGGAGCCCAATTCGAAATCTGAGGGTCAGAG 529  
|||||  
QY 1558 GAATCTTTCTGCTCCACAGAGCTCAGTGAACCCACAGAGGAGAGAGTGAACAAAGCT 1617  
|||||  
Db 528 GAATCTTTCTGCTCCACAGAGCTCAGTGAACCCACAGAGGAGAGAGTGAACAAAGCT 469  
|||||  
QY 1618 CCTGAAGAATCATCAGAGGATGTGACAAAATATCAGGAAGGAGTATCTGCAGAAACCCA 1677  
|||||  
Db 468 CCTGAAGAATCATCAGAGGATGTGACAAAATATCAGGAAGGAGTATCTGCAGAAACCCA 409  
|||||  
QY 1678 GTTGAGAACCATATCAATATACACAAATCAGATAAGTTTCAGACCAAGCCATTTGATTC 1737  
|||||  
Db 408 GTTGAGAACCATATCAATATACACAAATCAGATAAGTTTCAGACCAAGCCATTTGATTC 349  
|||||  
QY 1738 AACTCAGGAGAAAGAAATGACCTCAATCTTGATGCTCTTGTGGGGTCCAGAAAGATCT 1797  
|||||  
Db 348 AACTCAGGAGAAAGAAATGACCTCAATCTTGATGCTCTTGTGGGGTCCAGAAAGATCT 289  
|||||  
QY 1798 GCTTCATCTGAAAAAGCCAAAGGAACAGAACTTCAGATCAGACTAGCAGAGTGTCT 1857  
|||||  
Db 288 GCTTCATCTGAAAAAGCCAAAGGAACAGAACTTCAGATCAGACTAGCAGAGTGTCT 229  
|||||  
QY 1858 ACCAATGAAATTAACACCAATCTGAGCTCAGTTCCTCCAAACAGAGCCACTGGGCTTCA 1917  
|||||  
Db 228 ACCAATGAAATTAACACCAATCTGAGCTCAGTTCCTCCAAACAGAGCCACTGGGCTTCA 169  
|||||  
QY 1918 GCTCATGAGAAACATCCACAGGAGTCTGCTCTTCAGGACACAGATGACAGTGTATGAT 1977  
|||||  
Db 168 GCTCATGAGAAACATCCACAGGAGTCTGCTCTTCAGGACACAGATGACAGTGTATGAT 109  
|||||  
QY 1978 GACCCAGTCTCTGATCCAGGTGCAAGGTATCGAGAGGACCTGGGTATAGAGCCTTGTCT 2037  
|||||  
Db 108 GACCCAGTCTCTGATCCAGGTGCAAGGTATCGAGAGGACCTGGGTATAGAGCCTTGTCT 49  
|||||  
QY 2038 GTTGCCCGTATTTCAGGAGTCTTCAGACGAGAAAGAAAGAAAGAA 2085  
|||||  
Db 48 GTTGCCCGTATTTCAGGAGTCTTCAGACGAGAAAGAAAGAAAGAA 1

RESULT 30

BQ747286

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BE747286 601580236F1 NTH\_MGC\_9 Homo sapiens cDNA clone IMAGE:3929120 5',  
mRNA sequence.  
BE747286 BE747286.1 GI:10161278

EST.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 634)

NIH-MGC <http://mgs.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: c9apbs-re@mail.nih.gov  
 Tissue Procurement: DCTD/DTF  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
 Plate: L10CM761 row: g column: 09  
 High quality sequence stop: 634.  
 Location/Qualifiers  
 1. 634  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="IMAGE:3929120"  
 /tissue\_type="adrenocarcinoma cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: ovary; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
 EcoRI; cDNA made by oligo-dr priming. Directionally  
 cloned into EcoRI/XhoI sites using the following 5'  
 adaptor: GGACGAG(G). Size-selected >500bp for average  
 insert size 1.8kb. Library constructed by Ling Hong in  
 the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies)."  
 BASE COUNT 198 a 139 c 144 g 153 t  
 ORIGIN

FEATURES  
 source

Query Match 24.6%; Score 634; DB 12; Length 634;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-168;  
 Matches 634; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 773 CTCGTGTACAGTGAAGATGGTCAAGAGATTCGTTAGTTACTCTTCAGATTACATAT 832  
 Db 1 CTCGTGTACAGTGAAGATGGTCAAGAGATTCGTTAGTTACTCTTCAGATTACATAT 60

QY 833 ATCTTTTTCACCCGAAAGATGATACAGACGAGAACTTAAACTCCTCTCGGAGAGA 892  
 Db 61 ATCTTTTTCACCCGAAAGATGATACAGACGAGAACTTAAACTCCTCTCGGAGAGA 120

QY 893 GAAGAGAAGATGGGACACACACAGTTAAGCGTTTGAGACTTCGTGGTGTGTCAG 952  
 Db 121 GAAGAGAAGATGGGACACACACAGTTAAGCGTTTGAGACTTCGTGGTGTGTCAG 180

QY 953 ATACTGGACCCAGACGAGCGGAGAGTGAACGAGACGAGATGGAGCAGAGTCCCA 1012  
 Db 181 ATACTGGACCCAGACGAGCGGAGAGTGAACGAGACGAGATGGAGCAGAGTCCCA 240

QY 1013 ATGTGTCATGATGCAGAGATGTCTGATATGTTATCAAGATGGTTTGAAGAAGCAAGTG 1072  
 Db 241 ATGTGTCATGATGCAGAGATGTCTGATATGTTATCAAGATGGTTTGAAGAAGCAAGTG 300

QY 1073 AGTTTGCACAAAGCAATAGAGGACGAGAGATCTGCACCCAGAGTGGAAACAAGTCAAT 1132  
 Db 301 AGTTTGCACAAAGCAATAGAGGACGAGAGATCTGCACCCAGAGTGGAAACAAGTCAAT 360

QY 1133 CAGATATTTCAACTCTTCCTACGCTCCCATCAAGTCTGATTTGGAAGTGAAGTCAACTG 1192  
 Db 361 CAGATATTTCAACTCTTCCTACGCTCCCATCAAGTCTGATTTGGAAGTGAAGTCAACTG 420

QY 1193 CAATGGAAGTATGATCTCCAGCTGAACAATTTCTTACGCCCTTCTACATCTCTACAATGT 1252  
 Db 421 CAATGGAAGTATGATCTCCAGCTGAACAATTTCTTACGCCCTTCTACATCTCTACAATGT 480

QY 1253 CAGCTCAGGCTCATCTGCACATCATCTCCACAGAAAGCCCTCATCTACCTCTTGTGAT 1312  
 Db 481 CAGCTCAGGCTCATCTGCACATCATCTCCACAGAAAGCCCTCATCTACCTCTTGTGAT 540

QY 1313 CTCTCCAGACAGTGAACAAGGAGTCTCTGTGAGGCATCTGGACACACACACATCATC 1372  
 Db 541 CTCTCCAGACAGTGAACAAGGAGTCTCTGTGAGGCATCTGGACACACACACATCATC 600

QY 1373 AGTCTGATACAAATATGAAAAGCTGAGCCCCAA 1406

Db 601 AGTCTGATACAAATATGAAAAGCTGAGCCCCAA 634

RESULT 31  
 BM790042

LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

BM790042 635 bp mRNA linear EST 05-MAR-2002  
 K-EST0069804 S21SNU520 Homo sapiens cDNA clone S21SNU520-9-A04 5',  
 mRNA sequence.  
 BM790042  
 BM790042.1 GI:19138274  
 EST.  
 human.

REFERENCE  
 AUTHORS

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 635)  
 Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,  
 Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and  
 Kim,Y.S.

TITLE  
 JOURNAL  
 COMMENT

21C Frontier Korean EST Project 2001  
 Unpublished (2002)  
 Contact: Kim YS  
 Genome Research Center  
 Korea Research Institute of Bioscience & Biotechnology  
 52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea  
 Tel: +82-42-860-4470  
 Fax: +82-42-860-4409  
 Email: yongsung@mail.kribb.re.kr  
 Plate: 9 row: A column: 04  
 High quality sequence stop: 635.  
 Location/Qualifiers  
 1. 635  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="S21SNU520-9-A04"  
 /sex="F"  
 /tissue\_type="Stomach"  
 /cell\_type="Floating aggregates"  
 /lab\_host="Top10F"  
 /note="Organ: Stomach; Vector: pTZ18RP1; Site\_1: EcoRI;  
 Site\_2: NotI; The poly (A)+ RNA was dephosphorylated with  
 bacterial alkaline phosphatase (BAP) and then decapped  
 with tobacco acid pyrophosphatase (TAP). The decapped  
 intact mRNA was ligated with DNA-RNA linker including EcoR  
 I site by treatment of T4 RNA ligase and the first strand  
 cDNA was synthesized from oligo dt-selected mRNA by  
 priming with dt-tailed vector. The dt-tailed vector was  
 adjusted to have about 60nt. The cDNA vector was  
 circularized with E. coli DNA ligase after digestion of  
 EcoRI which site is also included in vector. An RNA strand  
 converted to a DNA strand by Okayama-Berg method. The  
 obtained cDNA vectors were used for transformation of  
 competent cells E. coli Top10F by electroporation method.  
 The cDNA libraries constructed by this method are  
 full-length enriched cDNA library."

FEATURES  
 source

BASE COUNT 218 a 148 c 144 g 125 t  
 ORIGIN

Query Match 24.6%; Score 633.4; DB 14; Length 635;  
 Best Local Similarity 99.8%; Pred. No. 4.3e-168;  
 Matches 634; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1435 TTGCACGTACAGCAGAGGAGCAACTACAGCACAATAAACTGAACCTTACAGATGAA 1494  
 Db 1 TTGCACGTACAGCAGAGGAGCAACTACAGCACAATAAACTGAACCTTACAGATGAA 60

QY 1495 TGGAGCAGTAGATCATCAAGTCTTAGAGGAATTTGGAGCCATTCGAATCTGGAGGTCAG 1554  
 Db 61 TGGAGCAGTAGATCATCAAGTCTTAGAGGAATTTGGAGCCATTCGAATCTGGAGGTCAG 120

QY 1555 GAGGAATCTTTCTGCCACAGAGCTCAGTGCACACAGAGGAGACAGTGAACAAAA 1614  
 Db 121 GAGGAATCTTTCTGCCACAGAGCTCAGTGCACACAGAGGAGACAGTGAACAAAA 180  
 QY 1615 GCTCCTGAAGAATCATCAGAGGATGTGACAAAATATATCAGGAAGAGATCTGCGAAAAAC 1674  
 Db 181 GCTCCTGAAGAATCATCAGAGGATGTGACAAAATATATCAGGAAGAGATCTGCGAAAAAC 240  
 QY 1675 CCAGTTGAGAACCATATCAATATATACCAATCAGATAAGTTTCACAGCAAGCCCAATTCGAT 1734  
 Db 241 CCAGTTGAGAACCATATCAATATATACCAATCAGATAAGTTTCACAGCAAGCCCAATTCGAT 300  
 QY 1735 TCCAACTCAGGAGAGAAAGAAATGACCTCAATCTTGATGCTCTTTGGGGTTTCCAGAGAA 1794  
 Db 301 TCCAACTCAGGAGAGAAAGAAATGACCTCAATCTTGATGCTCTTTGGGGTTTCCAGAGAA 360  
 QY 1795 TCTGCTTCATCTGAAAAGCCAAAGAACCAAGAACCTTCCAGATCAGACTAGCAGT 1854  
 Db 361 TCTGCTTCATCTGAAAAGCCAAAGAACCAAGAACCTTCCAGATCAGACTAGCAGT 420  
 QY 1855 GCTACCAATGAAATATACCAATCTTGAGCTCAGTTCCAAACAGAGCCACTGGGCT 1914  
 Db 421 GCTACCAATGAAATATACCAATCTTGAGCTCAGTTCCAAACAGAGCCACTGGGCT 480  
 QY 1915 TCAGCTCATGAAGAAACATCCACAGGAGCTCTGCTCTTCAGGACACAGATGACAGT 1974  
 Db 481 TCAGCTCATGAAGAAACATCCACAGGAGCTCTGCTCTTCAGGACACAGATGACAGT 540  
 QY 1975 GATGACCCAGTCCGATCCAGGTCAGAGTATCAGCAGGACCTGATAGAGGCTCT 2034  
 Db 541 GATGACCCAGTCCGATCCAGGTCAGAGTATCAGCAGGACCTGATAGAGGCTCT 600  
 QY 2035 GCTGTTGCCCTTATTCAGGAGTCTTTCAGAGGGAG 2069  
 Db 601 GCTGTTGCCCTTATTCAGGAGTCTTTCAGAGGGAG 635

RESULT 32  
 BG719267 707 bp mRNA linear EST 08-MAY-2001  
 LOCUS 602690335f1 NIH\_MGC\_97 Homo sapiens cDNA clone IMAGE:4822641 5',  
 DEFINITION mRNA sequence.

BG719267  
 VERSION BG719267.1 GI:13998454  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 707)  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki  
 Toshiyuki and Piero Carninci (RIKEN)  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov

Plate: L14M10731 row: e column: 10  
 High quality sequence stop: 704.  
 Location/Qualifiers

FEATURES  
 source

1. .707  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4822641"  
 /clone\_lib="NIH\_MGC\_97"  
 /lab\_host="DH10B"  
 /note="Organ: testis; Vector: pBluescript (modified)

pBluescript KS+); Site\_1: BamHI; Site\_2: SalI-XhoI (gtcgag  
 ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTNN-3',  
 size-selected for average insert size 2.2 kb and  
 normalized to ROT 5. This is a primary library enriched  
 for full-length clones and constructed using the  
 Cap-trapper method (Carninci, in preparation). Library  
 constructed by M. Brownstein (NIH/NHGRI, National  
 Institutes of Health). Note: this is a NIH\_MGC Library."

BASE COUNT 210 a 139 c 151 g 207 t  
 ORIGIN

Query Match 24.58; Score 631.4; DB 12; Length 707;  
 Best Local Similarity 96.1%; Pred. No. 1.7e-167;  
 Matches 658; Conservative 0; Mismatches 26; Indels 1; Gaps 1;

QY 210 AGATGACACCAATAGTAATTAATTCCTTACAGCAGAAAGGTTTTCACAACTTCG 269  
 Db 23 AGAGCACTGAGAAATATCTTTTAAGGAGAGGTTTCTCATCAGAGGTTTTCACAACTTCG 82  
 QY 270 TTCAGGGCACCAGAGAAACATATTTAGTGCAAGTTCTTACCTTGCTACAAATGATAACA 329  
 Db 83 TTCAGGGCACCAGAGCAACATATTTAGTGCAAGTTCTTACCTTGCTACAAATGATAACA 142  
 QY 330 GATTGTATCTGCTGAGATGGAGTAATATTTATACCAAGTTTGACCAAGATGCAGA 389  
 Db 143 GATTGTATCTGCTGAGATGGAGTAATATTTATACCAAGTTTGACCAAGATGCAGA 202  
 QY 390 AACCAACAGACAATGCCAATTTACGTGTCTATATGGAATCTTATGAGATTTACGTGT 449  
 Db 203 AACCAACAGACAATGCCAATTTACGTGTCTATATGGAATCTTATGAGATTTACGTGT 262  
 QY 450 ACCCAATGACCTTACACTTTCTCTCTGTTGTTGGAAGTGAAGTGTAGTGTTGA 509  
 Db 263 ACCCAATGACCTTACACTTTCTCTCTGTTGTTGGAAGTGAAGTGTAGTGTTGA 322  
 QY 510 TACAGCATCAAACTAGCTGCACAAAAGAGATTGTAAGATGATATTTTAACTG 569  
 Db 323 TACAGCATCAAACTAGCTGCACAAAAGAGATTGTAAGATGATATTTTAACTG 382  
 QY 570 TCGAGTGTGCCAGTCTGTGCTATTTGCCCAACCAATACCATATACCTTCTGCTGTTGG 629  
 Db 383 TCGAGTGTGCCAGTCTGTGCTATTTGCCCAACCAATACCATATACCTTCTGCTGTTGG 442  
 QY 630 TTGTTCTGACAGCTCAGTACGAATATATGATCGGGAATGCTGGGCACAAAGAGCTACAGG 689  
 Db 443 TTGTTCTGACAGCTCAGTACGAATATATGATCGGGAATGCTGGGCACAAAGAGCTACAGG 502  
 QY 690 GAATTATGAGTGCAGGAGCTACTGGAATGTTGCCCTTTTATCTTCCATCTTAA 749  
 Db 503 GAATTATGAGTGCAGGAGCTACTGGAATGTTGCCCTTTTATCTTCCATCTTAA 562  
 QY 750 TAATAAGTCTTCAGATGACATCTCTGTTACAGTGAAGATGGTCAAGAGATTTCTCGT 809  
 Db 563 TAATAAGTCTTCAGATGACATCTCTGTTACAGTGAAGATGGTCAAGAGATTTCTCGT 622  
 QY 810 TAGTTACTCTTCAGATGACATATCTTTTACCCGGAAGATGATACAGCAGGAGAACT 869  
 Db 623 TAGTTACTCTTCAGATGACATATCTTTTACCCGGAAGATGATACAGCAGGAGAACT 681  
 QY 870 TAAACTCTTCTCGGGAAGAGAGA 894  
 Db 682 TAAACTCTTCTCGGGAAGAGAGA 706

RESULT 33  
 BG534181

LOCUS 602553301f1 NIH\_MGC\_77 Homo sapiens cDNA clone IMAGE:462816 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BG534181  
 VERSION BG534181.1 GI:13525721  
 KEYWORDS EST.  
 SOURCE human.

BG534181 728 bp mRNA linear EST 03-APR-2001  
 602553301f1 NIH\_MGC\_77 Homo sapiens cDNA clone IMAGE:462816 5',  
 mRNA sequence.

BG534181  
 BG534181.1 GI:13525721  
 EST.  
 human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 728)  
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: CLONTECH Laboratories, Inc.  
 cDNA Library Preparation: CLONTECH Laboratories, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: L1CM1464 row: b column: 01  
 High quality sequence stop: 681.  
 FEATURES  
 Location/Qualifiers  
 1..728  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4662816"  
 /clone.lib="NIH\_MGC\_77"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: lung; Vector: pDMR-LIB (Clontech); Site: 1:  
 SfiI (ggccgctggccc); Site: 2: SfiI (ggccattggccc); 5' and  
 3' adaptors were used in cloning as follows: 5' adaptor  
 sequence: 5'-CAGGCGCATATGCCC-3' and 3' adaptor sequence:  
 5'-ATTCTAGAGCGCGGCGGCAGACATG-dt(30)BN-3' (where B = A,  
 C, or G and N = A, C, G, or T). Average insert size 1.9  
 kb (range 0.5-4.0 kb). 12/15 colonies contained inserts  
 by PCR. This library was enriched for full-length clones  
 and was constructed by Clontech Laboratories (Palo Alto,  
 CA). Note: this is a NIH\_MGC Library."  
 BASE COUNT 221 a 153 c 175 g 179 t  
 ORIGIN  
 Query Match 24.4%; Score 629.8; DB 12; Length 728;  
 Best Local Similarity 96.7%; Pred. No. 4.9e-167;  
 Matches 707; Conservative 0; Mismatches 17; Indels 7; Gaps 6;  
 QY 663 GCGAATGCTGGGCAACAGCTACAGGAATTATTCAGGTCGAGGACTACATGGAATGTT 722  
 Db 1  
 QY 2 GCGAATGCTGGGCAACAGCTACAGGAATTATTCAGGTCGAGGACTACATGGAATGTT 61  
 Db 1  
 QY 723 TGCCCGTTTATTCCTCCCATCTTAATAAATAGTCCTGCAGAGTGACATCTCTGTGTTA 782  
 Db 62 TGCCCGTTTATTCCTCCCATCTTAATAAATAGTCCTGCAGAGTGACATCTCTGTGTTA 121  
 QY 783 CAGTGAAGATGGTCAAGAGATTCCTGTAGTTACTCTTCAGATTCACATATATCTTTTGA 842  
 Db 122 CAGTGAAGATGGTCAAGAGATTCCTGTAGTTACTCTTCAGATTCACATATATCTTTTGA 181  
 QY 843 CCGCAAGATGATACAGCAGAGAACTTAAACCTCTTCGCGAAGAGAGAGAGAAGA 902  
 Db 182 CCGCAAGATGATACAGCAGAGAACTTAAACCTCTTCGCGAAGAGAGAGAGAAGA 241  
 QY 903 GTTCCGACACACCACCTTAAGCGTTTGAGACTTCGTGGTGGTTCAGATCTGGACC 962  
 Db 242 G-TCCGACACACCACCTTAAGCG-TTGAGACTTCGTGGTGGTTCAGATCTGGACC 298  
 QY 963 CAGAGCAAGCGCGGAGAGTGAACGAGAACGAGATGGAGAGAGAGTCCCAATGTGTCATT 1022  
 Db 299 CAGAGCAAGCGCGGAGAGTGAACGAGAACGAGATGGAGAGAGAGTCCCAATGTGTCATT 358  
 QY 1023 GATGCGAGAGATGTCTGATATGTTTATCAAGATGTTTGAAGCAAGTGAAGTTGCACA 1082  
 Db 359 GATGCGAGAGATGTCTGATATGTTTATCAAGATG-TTGAAGCAAGTGAAGTTGCACA 417  
 QY 1083 AAGCAATAGGAGGAGAGATCTGACCCAGAGGTGGAAACAAGTCAATCAGATATTTTC 1142  
 Db 418 AAGCAATAGGAGGAGAGATCTGACCCAGAGGTGGAAACAAGTCAATCAGATATTTTC 477

QY 1143 AACTCTTCTACGTCCTCCATCAAGTCCCTGATTTGGAGTGAGTGAAGTCAATGGAAGT 1202  
 Db 478 AACTCTTCTACGTCCTCCATCAAGTCCCTGATTTGGAGTGAGTGAAGTCAATGGAAGT 537  
 QY 1203 AGATACTCCAGCTGAACAATTTCTTCAGCTTCTACATCTCTACATATGTCACGTCAGGC 1262  
 Db 538 AGATACTCCAGCTGAACAATTTCTTCAGCTTCTACATCTCTACATATGTCACGTCAGGC 597  
 QY 1263 TCATTTCGACATCATCTCCACAGAAGCCCTCATCTTCTCTCTCTCTCTCTCTCTCT 1322  
 Db 598 TCATTTCGA-ATCATCTCCACAGAAGCCCTCATCTTCTCTCTCTCTCTCTCTCTCT 656  
 QY 1323 CAGTGAACAAGGCGAGCTGTGTTGAG--CATCTGGACACACACATCATCATGTCGAT 1380  
 Db 657 TAGTGAACAAGGCGAGCTGTGTTGAGGCGATCTGGAACACCAACCAAACTTCATCAGTCTG 716  
 QY 1381 AACAAATATGA 1391  
 Db 717 AACAAATATGA 727  
 RESULT 34  
 BI459472  
 LOCUS  
 DEFINITION 869 bp mRNA linear EST 21-AUG-2001  
 603200466F1 NIH\_MGC\_97 Homo sapiens cDNA clone IMAGE:5266144 5',  
 mRNA sequence.  
 ACCESSION BI459472  
 VERSION BI459472.1 GI:15250128  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 869)  
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki  
 Toshiyuki and Piero Carninci (RIKEN)  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: L1AM11671 row: d column: 17  
 High quality sequence stop: 732.  
 FEATURES  
 Location/Qualifiers  
 1..869  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5266144"  
 /clone.lib="NIH\_MGC\_97"  
 /lab\_host="DH10B"  
 /note="Organ: testis; Vector: pBluescriptR (modified  
 pBluescript KS+); Site: 1: BamHI; Site: 2: SalI-XhoI (gtcgag  
 ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTTVN-3',  
 size-selected for average insert size 2.2 kb and  
 normalized to ROT 5. This is a primary library enriched  
 for full-length clones and constructed using the  
 Cap-trapper method (Carninci, in preparation). Library  
 constructed by M. Brownstein (NIH/NHGRI, National  
 Institutes of Health). Note: this is a NIH\_MGC Library."  
 BASE COUNT 257 a 163 c 214 g 235 t  
 ORIGIN  
 Query Match 24.4%; Score 628.6; DB 13; Length 869;  
 Best Local Similarity 92.5%; Pred. No. 1.2e-166;  
 Matches 750; Conservative 0; Mismatches 49; Indels 12; Gaps 8;









); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.2 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH\_MGC Library."

BASE COUNT 222 a 144 c 156 g 211 t  
ORIGIN

Query Match 24.2%; Score 624.4; DB 13; Length 733;  
Best Local Similarity 95.9%; Pred. No. 1.7e-165;  
Matches 673; Conservative 0; Mismatches 26; Indels 3; Gaps 3;

```

QY 220 AAATAGTATTAGTAACTCTTACAGCAGAAAGGTTTGGACAACTTCGTTACGGGCAC 279
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 32 AAATAGTATTAGTAAAGAGGTTTCTCATCAGAGGTTTGGACAACTTCGTTACGGGCAC 91
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 280 CGAGCAACATATTTAGTGCACAAAGTCTTACCTTGTACAAATGATAACACAGATTGTATCC 339
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 92 CGAGCAACATATTTAGTGCACAAAGTCTTACCTTGTACAAATGATAACACAGATTGTATCC 151
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 340 TGTCTGGAGATGAGTAATATTTATACCAAGCTTGAGCAAGATGCAAGAACCAACAGAGA 399
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 152 TGTCTGGAGATGAGTAATATTTATACCAAGCTTGAGCAAGATGCAAGAACCAACAGAGA 211
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 400 CAATGCCAATTTACGTGTCATTTATGGAATCTATTATGAGATTATGACTGTACCCCAATGAC 459
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 212 CAATGCCAATTTACGTGTCATTTATGGAATCTATTATGAGATTATGACTGTACCCCAATGAC 271
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 460 CCTTACACTTTTCTCTCTGTTGGTGAAGATGGAATCTTTAGTGGTTTGTATACAGCAATC 519
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 272 CCTTACACTTTTCTCTCTGTTGGTGAAGATGGAATCTTTAGTGGTTTGTATACAGCAATC 331
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 520 AAACTAGCTGCACAAAAGATTTGTAAGATGATATTTTAATTAACCTGTGCGAGTGTCT 579
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 332 AAACTAGCTGCACAAAAGATTTGTAAGATGATATTTTAATTAACCTGTGCGAGTGTCT 391
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 580 GCCAGCTGTGTTGTTATTTGCCACCAATACCAATATACCTTGTGTTGTTGTTCTGAC 639
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 392 GCCAGCTGTGTTGTTATTTGCCACCAATACCAATATACCTTGTGTTGTTGTTCTGAC 451
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 640 AGCTCAGTACGAATATATGTCGGGAATGTCGGCAAGAGCTACAGGGAATTTATGCA 699
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 452 AGCTCAGTACGAATATATGTCGGGAATGTCGGCAAGAGCTACAGGGAATTTATGCA 511
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 700 GGTCTGAGGACTACTGGAATGTTGCCGTTTATTCCTTCCCATCTTAATAAATAGTCC 759
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 512 GGTCTGAGGACTACTGGAATGTTGCCGTTTATTCCTTCCCATCTTAATAAATAGTCC 571
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 760 TGCAGAGTGACATCTCTGTGTTACAGTGAAGATGTCGAAGATTCCTGTTAGTTACTCT 819
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 572 TGCAGAGTGACATCTCTGTGTTACAGTGAAGATGTCGAAGATTCCTGTTAGTTACTCT 631
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 820 TCAGATTACATATATCTTTTTCACCG - AAACATGATACAGCAGCAGAC - TTAACACTC 877
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 632 TCAGATTACATATATCTTTTTCACCGAAGATGATACAGCAGCAGACACTTAAACACTC 691
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 878 CTTCTCGGGAAGAG - AGAAGAAGAGTTTGCACACACCA 918
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 692 CTTCTCGGGAAGAGAGAGAGAGAGTTGCGACACCA 733
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

RESULT 38  
BM926145  
LOCUS BM926145 1199 bp mRNA linear EST 12-MAR-2002  
DEFINITION AGENCOURT\_6600775 NIH\_MGC\_114 Homo sapiens cdna clone IMAGE:5764509  
5', mRNA sequence.  
ACCESSION BM926145  
VERSION BM926145.1 GI:19376524  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

# REFERENCE AUTHORS TITLE JOURNAL COMMENT

1 (bases 1 to 1199)  
NIH-MGC <http://mhc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)  
Tissue Procurement: Life Technologies, Inc.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLML)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLML at:  
<http://image.llnl.gov>

Plate: LLML2818 row: a column: 22  
High quality sequence start: 28  
High quality sequence stop: 546.  
Location/Qualifiers  
1. 1199

## FEATURES source

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5764509"  
/clone\_lib="NIH\_MGC\_114"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: EcoRV (destroyed); RNA source anonymous pool of 6  
male brains, age range 23-27 yo. Library is oligo-dT  
primed and directionally cloned (EcoRV site is destroyed  
upon cloning). Average insert size 1.5 kb, insert size  
range 1-3 kb. Library is normalized and enriched for  
full-length clones and was constructed by C. Gruber  
(Invitrogen). Research Genetics tracking code 019. Note:  
this is a NIH\_MGC Library."

BASE COUNT 359 a 290 c 251 g 292 t 7 others  
ORIGIN

Query Match 24.2%; Score 624; DB 14; Length 1199;  
Best Local Similarity 96.4%; Pred. No. 2.8e-165;  
Matches 692; Conservative 0; Mismatches 20; Indels 6; Gaps 5;

```

QY 44 GGAAGAAGTCCCTGGGCTGGAGAGCCGCTCCGCTCGGAGTCCGCTACCTGGGAAGAA 103
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 5 GCAAAAGGGGGCTGGGCTGGAGGA - ATGCCCGGCTGGGAGTCCGCTACCTGGGAAGAA 63
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 104 GAGATTTATCAAGATTAAACTTGAGCAACCTTAATGTCATGATGTTGTGTTA 163
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 64 GAGATTTATCAAGATTATATCTTGAAGCAACCCCTTAATGTCATGATGTTGTGTTA 123
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 164 ATACAATCTGTTGGAGTACACT - GGAGAAATATATTTATCTGGCTCAGATGACACCAA 222
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 124 ATACAATCTGTTGGAGTACACTGGGAGAAATATTTATCTGGCTCAGATGACACCAA 183
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 223 TTAGTAATTTAGTAATCTTTACAGCAAGAGTTTGGACAACTTCGTTACGGGCACCA 282
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 184 TTAGTAATTTAGTAATCTTTACAGCAAGAGTTTGGACAACTTCGTTACGGGCACCA 243
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 283 GCAACATATTTAGTGCACAAAGTCTTACCTTGTACAAATGATAACAGATTGTATCTGC 342
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 244 GCAACATATTTAGTGCACAAAGTCTTACCTTGTACAAATGATAACAGATTGTATCTGC 303
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 343 TCTGGAGATGGAGTAATATTTATACCAACGTTGACAAAGATGCAAGAACCAACAGACAA 402
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 304 TCTGGAGATGGAGTAATATTTATACCAACGTTGACAAAGATGCAAGAACCAACAGACAA 363
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 403 TGCATTTACGTGTCATTTATGGAATCTATGAGATTATGACTGTACCAATGACCT 462
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 364 TGCATTTACGTGTCATTTATGGAATCTATGAGATTATGACTGTACCAATGACCT 423
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 463 TACACTTTCTCTCTGTTGGTGAAGATGGAACCTGTTAGTGGTTTGTATACACCATCAA 522
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 424 TACACTTTCTCTCTGTTGGTGAAGATGGAACCTGTTAGTGGTTTGTATACACCATCAA 483
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```



found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
plate: L1CM849 row: o column: 22  
High quality sequence stop: 587.

## FEATURES

## source

1. .912  
/location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3996909"  
/clone\_lib="NIH\_MGC\_53"  
/tissue\_type="carcinoma, cell line"  
/lab\_host="DH10B (TI phage-resistant)"  
/note="Organ: bladder; Vector: pDNR-LIB (Clontech);  
Site\_1: SfII (ggcgctcgcc); Site\_2: SfII (ggccattggcc  
); Double-stranded cDNA was prepared from cell line RNA.  
5' and 3' adaptors were used in cloning as follows: 5'  
adaptor sequence: 5'-ATTCTAGAGCGGCGCCGACATG-dT(30)BN-3'  
sequence: 5'-ATTCTAGAGCGGCGCCGACATG-dT(30)BN-3'  
(where B = A, C, or G and N = A, C, G, or T). Average  
insert size 1.55 kb (range 0.9-4.0 kb). 15/15 colonies  
contained inserts by PCR. This library was enriched for  
full-length clones and was constructed by Clontech  
Laboratories (Palo Alto, CA)."

BASE COUNT 294 a 216 c 228 g 174 t  
ORIGIN

Query Match 24.0%; Score 618.4; DB 12; Length 912;  
Best Local Similarity 97.1%; Pred. No. 9.3e-164;  
Matches 662; Conservative 0; Mismatches 16; Indels 4; Gaps 3;

Qy 1575 GAGCTCAGTGAACACACAGAGGAGAGTGTCTCGAAGAACCCAGTTGAGAACCATATCAA 1634  
Db 1 GAGCTCAGTGAACACACAGAGGAGAGTGTCTCGAAGAACCCAGTTGAGAACCATATCAA 60  
Qy 1635 GGATGTGACAAATATCAGGAGGAGTGTCTCGAAGAACCCAGTTGAGAACCATATCAA 1694  
Db 61 GGATGTGACAAATATCAGGAGGAGTGTCTCGAAGAACCCAGTTGAGAACCATATCAA 120  
Qy 1695 TATAACACAATCAGATAAGTTTACAGCCAGCCATTTGGATTCCAACTCAGGAGAGAGAA 1754  
Db 121 TATAACACAATCAGATAAGTTTACAGCCAGCCATTTGGATTCCAACTCAGGAGAGAGAA 180  
Qy 1755 TGACCTCAATCTTGATCGCTCTGTGGGTTCAGAGAAATCTGCTCATCTGAAAGAGC 1814  
Db 181 TGACCTCAATCTTGATCGCTCTGTGGGTTCAGAGAAATCTGCTCATCTGAAAGAGC 240  
Qy 1815 CAAGGAACAGAACTTCAGATCAGACTAGCAGTGTGAGAGTGTACCAATGAAATTAACAC 1874  
Db 241 CAAGGAACAGAACTTCAGATCAGACTAGCAGTGTGAGAGTGTACCAATGAAATTAACAC 300  
Qy 1875 CAATCTGAGCCTCAGTTCCAAACAGAGCCACTGGGCTTCAGCTCATGAGAGAAACATC 1934  
Db 301 CAATCTGAGCCTCAGTTCCAAACAGAGCCACTGGGCTTCAGCTCATGAGAGAAACATC 360  
Qy 1935 CACGAGGAGCTGCTCTTCAGAGACACATCAGATGATGATGATGATGATGATGATGATGAT 1994  
Db 361 CACGAGGAGCTGCTCTTCAGAGACACATCAGATGATGATGATGATGATGATGATGATGAT 420  
Qy 1995 AGGTGCAAGGTATCGAGCAGGAGCTGGTGTATAGACGCTCTGCTGTGTGCGCGTATTCAGGA 2054  
Db 421 AGGTGCAAGGTATCGAGCAGGAGCTGGTGTATAGACGCTCTGCTGTGTGCGCGTATTCAGGA 480  
Qy 2055 GTTCTTCAGAGGAG 2114  
Db 481 GTTCTTCAGAGGAG 540  
Qy 2115 AAGGCGGTAGTAAATGTTTATAAAGGCGCATCCCAACTCCAGGACAAATGATAAAGA 2174  
Db 541 AAGGCGG-TAGTAAATGTTTATAA--GGCATCCCAACTCCAGGAC-ATGATGAAGA 596  
Qy 2175 AGCCAATTTCTGGGGTCTAACTTTTAAATAGTGTGTTCTGACTGTGGCCACATTTTCAT 2234  
Db 597 AGCCAATTTCTGGGGTCTCACCTTGGAAATGAGTGTGCTGACTGTGGACACATTTTCAT 656

Qy 2235 CTGGGATCGGCACACTGCTGAG 2256  
||||||| ||||| ||  
Db 657 CTGGGATCGGAACACTGTGAAG 678

Search completed: March 26, 2003, 03:07:25  
Job time : 4101 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2003, 20:23:58 ; Search time 104 seconds

(without alignments)

7607.943 Million cell updates/sec

Title: US-09-781-693A-1

Perfect score: 2580

Sequence: 1 atgtctgggggtgctcta.....atgaaatgagatgagaa 2580

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 120 summaries

Database :

Issued\_Patents\_NA:\*

1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*

2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*

3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq:\*

4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq:\*

5: /cgn2\_6/ptodata/1/ina/PCTUS.COMB.seq:\*

6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	59	2.3	7218	1	US-08-232-463-14
2	39	1.5	429	4	US-08-847-065-20
3	39	1.5	2462	4	US-08-847-065-24
4	39	1.5	3023	4	US-09-308-022-4
5	38.4	1.5	2085	1	US-08-072-070-1
6	38.4	1.5	2085	1	US-08-465-746-1
7	38.4	1.5	2085	1	US-08-214-164-1
8	38.4	1.5	2085	1	US-08-469-434-1
9	38.4	1.5	2085	1	US-08-214-222-1
10	38.4	1.5	2085	2	US-08-467-852A-1
11	38.4	1.5	2085	2	US-08-468-718-1
12	38.4	1.5	2085	2	US-08-246-636-1
13	38.4	1.5	2085	2	US-08-247-494A-1
14	38.4	1.5	2085	2	US-08-319-795-1
15	38.4	1.5	2085	2	US-08-468-985-1
16	38.4	1.5	2086	3	US-08-312-949-1
17	38.4	1.5	2086	3	US-08-446-201-2
18	37.4	1.4	1360	3	US-08-961-083-37
19	37.4	1.4	12665	4	US-08-961-527-134
20	35.8	1.4	289	4	US-09-007-005-17
21	35.8	1.4	289	4	US-09-244-796-17
22	35.6	1.4	1695	4	US-09-308-022-5
23	35.4	1.4	598	4	US-09-385-982-267
C 24	35.4	1.4	1188	3	US-08-048-700-1
C 25	35.4	1.4	1188	3	US-08-839-711-5
C 26	34.6	1.3	405	2	US-08-299-074A-1
C 27	34.6	1.3	405	4	US-09-399-773-1

C 28	34.6	1.3	1815	2	US-08-606-288-6	Sequence 6, Appli
C 29	34.6	1.3	1815	3	US-09-347-483-6	Sequence 6, Appli
30	34.4	1.3	248	4	US-09-007-005-32	Sequence 32, Appli
31	34.4	1.3	248	4	US-09-244-796-32	Sequence 32, Appli
32	34.4	1.3	277	4	US-09-007-005-3	Sequence 3, Appli
33	34.4	1.3	277	4	US-09-244-796-3	Sequence 3, Appli
34	34.4	1.3	913	1	US-08-217-327-3	Sequence 3, Appli
35	34.4	1.3	913	1	US-08-885-970A-3	Sequence 3, Appli
36	34.4	1.3	913	1	US-08-298-687A-3	Sequence 3, Appli
37	34.4	1.3	913	1	US-08-530-797-2	Sequence 3, Appli
38	34.4	1.3	913	1	US-08-298-829-3	Sequence 3, Appli
39	34.4	1.3	913	2	US-08-787-335-2	Sequence 2, Appli
40	34.4	1.3	1984	1	US-07-885-970A-25	Sequence 2, Appli
41	34.4	1.3	1985	1	US-08-298-687A-25	Sequence 25, Appli
42	34.4	1.3	1985	1	US-08-298-829-25	Sequence 25, Appli
43	33.6	1.3	7100	4	US-09-308-375-1	Sequence 1, Appli
44	33.4	1.3	825	4	US-09-134-001C-2307	Sequence 1084, Ap
45	33.4	1.3	924	4	US-09-134-001C-2307	Sequence 2307, Ap
46	33.2	1.3	282	4	US-09-461-697-205	Sequence 205, App
47	33.2	1.3	306	4	US-09-461-697-203	Sequence 193, App
48	33.2	1.3	696	4	US-09-461-697-193	Sequence 191, App
49	33.2	1.3	699	4	US-09-461-697-191	Sequence 189, App
50	33.2	1.3	717	4	US-09-461-697-189	Sequence 187, App
51	33.2	1.3	774	4	US-09-461-697-187	Sequence 185, App
52	33.2	1.3	819	4	US-09-461-697-185	Sequence 183, App
C 53	33.2	1.3	957	4	US-09-123-912-113	Sequence 184, App
C 54	33.2	1.3	957	4	US-09-643-597-113	Sequence 3, Appli
C 55	33.2	1.3	1669	4	US-09-461-697-184	Sequence 3, Appli
C 56	33.2	1.3	4296	4	US-09-060-410-3	Sequence 3, Appli
C 57	33	1.3	5433	3	US-08-929-329-1	Sequence 1, Appli
C 58	33	1.3	10607	1	US-08-078-080-3	Sequence 1, Appli
C 59	32.8	1.3	4617	4	US-08-930-055A-1	Sequence 1, Appli
60	32.8	1.3	5109	4	US-08-930-055A-2	Sequence 2, Appli
61	32.8	1.3	5567	4	US-08-899-241-1	Sequence 1, Appli
62	32.6	1.3	1683	4	US-09-347-803-11	Sequence 11, Appli
63	32.6	1.3	2652	4	US-09-115-704-1	Sequence 1, Appli
C 64	32.6	1.3	12666	4	US-08-961-527-137	Sequence 137, App
C 65	32.4	1.3	444	4	US-07-757-022B-37	Sequence 37, Appl
C 66	32.4	1.3	567	4	US-07-757-022B-31	Sequence 31, Appl
C 67	32.4	1.3	1665	4	US-09-134-001C-1565	Sequence 1565, Ap
C 68	32.4	1.3	13865	3	US-09-009-217-11	Sequence 11, Appl
C 69	32.4	1.3	13865	3	US-09-009-656-11	Sequence 11, Appl
C 70	32.4	1.3	31491	4	US-09-360-186-1	Sequence 1, Appli
C 71	32.2	1.2	645	4	US-09-328-111-9	Sequence 9, Appli
C 72	32.2	1.2	822	3	US-08-937-271-5	Sequence 5, Appli
C 73	32.2	1.2	5893	1	US-08-582-126-54	Sequence 54, Appl
C 74	32.2	1.2	12720	1	US-08-403-866-11	Sequence 11, Appl
C 75	32.2	1.2	29604	3	US-08-781-891-207	Sequence 207, App
C 76	32	1.2	624	4	US-09-397-992A-3	Sequence 3, Appli
C 77	32	1.2	624	4	US-09-397-992A-6	Sequence 6, Appli
C 78	31.8	1.2	497	3	US-09-284-782-21	Sequence 21, Appl
C 79	31.8	1.2	606	4	US-09-134-001C-747	Sequence 747, App
C 80	31.8	1.2	921	4	US-09-072-596-298	Sequence 298, App
C 81	31.8	1.2	1062	4	US-09-134-001C-278	Sequence 278, App
C 82	31.8	1.2	2799	1	US-08-212-188-1	Sequence 1, Appli
C 83	31.8	1.2	2799	3	US-08-970-725-1	Sequence 1, Appli
C 84	31.8	1.2	2859	5	PCT-US96-052708-1	Sequence 1, Appli
C 85	31.8	1.2	2859	5	PCT-US96-052708-1	Sequence 637, App
C 86	31.8	1.2	3522	1	US-08-940-751-4	Sequence 4, Appli
C 87	31.8	1.2	3522	1	US-08-291-568-1	Sequence 1, Appli
C 88	31.8	1.2	3522	2	US-08-962-190-1	Sequence 1, Appli
C 89	31.8	1.2	3522	2	PCT-US95-10310-1	Sequence 1, Appli
C 90	31.8	1.2	3522	6	5164180-3	Patent No. 5164180
C 91	31.6	1.2	2952	4	US-09-318-794A-4	Sequence 4, Appli
C 92	31.6	1.2	3498	3	US-08-293-728-1	Sequence 1, Appli
C 93	31.6	1.2	3498	4	US-09-421-868-1	Sequence 1, Appli
C 94	31.6	1.2	3666	2	US-08-682-517-13	Sequence 13, Appl
C 95	31.6	1.2	3666	2	US-08-682-517-14	Sequence 14, Appl
C 96	31.6	1.2	4197	2	US-08-682-517-7	Sequence 7, Appli
C 97	31.6	1.2	4197	2	US-08-682-517-8	Sequence 8, Appli
C 98	31.6	1.2	4378	2	US-09-080-897-3	Sequence 3, Appli
C 99	31.6	1.2	4378	2	US-09-323-735-3	Sequence 2, Appli
100	31.6	1.2	4399	3	US-08-899-595-2	



; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: CDNA  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: S. pneumoniae  
 US-08-847-065-20

Query Match 1.5%; Score 39; DB 4; Length 429;  
 Best Local Similarity 45.1%; Pred No. 0.053;  
 Matches 144; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

QY 1611 AAAAGCTCCTGAAGAACATCATCAGAGGATGTGACAAAATATCAGGAAGGAGTATCTGCAGA 1670  
 DB 1 AAAAAGGTAGCAGAAGCTGAGAAGAGTTGAAGAAGCTGAGAAAAAGCCAAAGGATCA 60  
 QY 1671 AAACCCAGTTGAGAACCATATCAATATACACAATCAGATAAGTTTACAGCCAGCCATT 1730  
 DB 61 AAAAGAAGAGATCCCGTGAATACCCCAACCAATATCTTACAAAACGCTTTGACCTTGAAT 120  
 QY 1731 GGATTCACCTCAGGAGAAAGAAATGACCTCAATCTTGATCGCTCTTGTTGGGTTCCAGA 1790  
 DB 121 TGCTGAGTCCGATGTGAAGTTTAAAGAGCGGAGCTTGAACCTAGTAAAGAGGAAGCTAA 180  
 QY 1791 AGAATCTGCTTCATCTGAAAAGCCCAAGAACCCAGAACTTCAGATCAGACTAGCACTGA 1850  
 DB 181 GGAACCTCGAGACGAGGAGGAAAAATTAAGCAAGCAAGAAAGCGGAAAGTTGAGAGTAA 240  
 QY 1851 GAGTGCTACCAATGAAATAACACCAATCCTGAGCTCAGTTCCAAACAGAGCCACTGG 1910  
 DB 241 TGAGCTACAGGTTAGAAACATCAGACAGATCGTAAAAAGCAGAGAAGAAAGCTAA 300  
 QY 1911 GCCTTCAGCTCATGAAGAA 1929  
 DB 301 ACGAAAGCAGCAGAAGAA 319

## RESULT 3

US-08-847-065-24

; Sequence 24, Application US/08847065  
 ; Patent No. 6245335  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Masure, H. Robert  
 ; APPLICANT: Rosenow, Carsten I.  
 ; APPLICANT: Tuomanen, Elaine  
 ; APPLICANT: Wizemann, Theresa M.  
 ; TITLE OF INVENTION: CHOLINE BINDING PROTEINS FOR  
 ; TITLE OF INVENTION: ANTI-PNEUMOCOCCAL VACCINES  
 ; NUMBER OF SEQUENCES: 25  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: David A. Jackson, Esq.  
 ; STREET: 411 Hackensack Ave, Continental Plaza, 4th  
 ; CITY: Hackensack  
 ; STATE: New Jersey  
 ; COUNTRY: USA  
 ; ZIP: 07601

; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/847,065  
 ; FILING DATE:

; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Jackson Esq., David A.  
 ; REGISTRATION NUMBER: 26,742  
 ; REFERENCE/DOCKET NUMBER: 600-1-158  
 ; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 201-487-5800  
 ; TELEFAX: 201-343-1684  
 ; INFORMATION FOR SEQ ID NO: 24:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2462 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: CDNA  
 ; HYPOTHETICAL: NO  
 ; US-08-847-065-24

Query Match 1.5%; Score 39; DB 4; Length 2462;  
 Best Local Similarity 45.1%; Pred No. 0.16; Indels 0; Gaps 0;  
 Matches 144; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

QY 1611 AAAAGCTCCTGAAGAACATCATCAGAGGATGTGACAAAATATCAGGAAGGAGTATCTGCAGA 1670  
 DB 768 AAAAAGGTAGCAGAAGCTGAGAAGAGTTTGAAGAAGCTGAGAAAAAGCCAAAGGATCA 827  
 QY 1671 AAACCCAGTTGAGAACCATATCAATATACACAATCAGATAAGTTTACAGCCAGCCATT 1730  
 DB 828 AAAAGAAGAGATCCCGTGAATACCCCAACCAATATCTTACAAAACGCTTTGACCTTGAAT 887  
 QY 1731 GGATTCACCTCAGGAGAAAGAAATGACCTCAATCTTGATCGCTCTTGTTGGGTTCCAGA 1790  
 DB 888 TGCTGAGTCCGATGTGAAGTTTAAAGAGCGGAGCTTGAACCTAGTAAAGAGGAAGCTAA 947  
 QY 1791 AGAATCTGCTTCATCTGAAAAGCCCAAGAACCCAGAACTTCAGATCAGACTAGCACTGA 1850  
 DB 948 GGAACCTCGAGACGAGGAGGAAAAATTAAGCAAGCAAGAAAGCGGAAAGTTGAGAGTAA 1007  
 QY 1851 GAGTGCTACCAATGAAATAACACCAATCCTGAGCTCAGTTCCAAACAGAGCCACTGG 1910  
 DB 1008 TGAGCTACAGGTTAGAAACATCAGACAGATCGTAAAAAGCAGAGAAGAAAGCTAA 1067  
 QY 1911 GCCTTCAGCTCATGAAGAA 1929  
 DB 1068 ACGAAAGCAGCAGAAGAA 1086

## RESULT 4

US-09-308-022-4

; Sequence 4, Application US/09308022  
 ; Patent No. 6291654  
 ; GENERAL INFORMATION:  
 ; APPLICANT: REGENTS OF THE UNIVERSITY OF MINNESOTA, ET AL.  
 ; TITLE OF INVENTION: C3 BINDING PROTEIN OF STREPTOCOCCUS  
 ; TITLE OF INVENTION: PNEUMONIAE  
 ; NUMBER OF SEQUENCES: 12  
 ; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MUETING, RAASCH & GEBHARDT, P.A.  
 ; STREET: 119 No. 6291654th Fourth Street, Suite 203  
 ; CITY: Minneapolis  
 ; STATE: Minnesota  
 ; COUNTRY: USA  
 ; ZIP: 55401

; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/308,022  
 ; FILING DATE:

; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US97/20586  
 ; FILING DATE: 12-NOV-1997  
 ; APPLICATION NUMBER: 60/029,444  
 ; FILING DATE: 12-NOV-1996  
 ; APPLICATION NUMBER: 60/038,086  
 ; FILING DATE: 18-FEB-1997

APPLICATION NUMBER: 60/059,368  
FILING DATE: 19-SEP-1997  
CLASSIFICATION: 60/062,473  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 60/062,473  
FILING DATE: 16-OCT-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: MUEING, Ann M.  
REGISTRATION NUMBER: 33,977  
REFERENCE/DOCKET NUMBER: 110.00430101  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 612-305-1217  
TELEFAX: 612-305-1228  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3023 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-09-308-022-4

Query Match 1.5%; Score 39; DB 4; Length 3023;  
Best Local Similarity 45.1%; Pred. No. 0.19;  
Matches 144; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

QY 1611 AAAAGCTCTGAGAAATCATCAGAGGATGTGACAAAATATCAGGAAGGATCTGCGA 1670  
DB 1282 AAAAAAGGTGACAGAGCTGAGAGAGGTTGAAGAGCTGAGAAAAAGCCAGGATCA 1341

QY 1671 AAACCCAGTTGAGAACCATATCAATATACACATCAATCAGATAGTTCCAGCCGCAAT 1730  
DB 1342 AAAAGAGAGATCGCCGTAATACCCCAACCAATACCTTACAAAAGCTTGACCTTGAAT 1401

QY 1731 GGATTCCAACTCAGGAGAAAGAAATGACCTCAATCTTGATCGCTCTTGTGGGTTCCAGA 1790  
DB 1402 TGCTGAGTCCGATGTGAAGTTAAAGAGCGGAGCTTCACTAGTAAAAGAGGAGCTAA 1461

QY 1791 AGAATCTGCTTCACTGTAAGAACGAGGAGAACCACTTCACTAGATAGTACACTGA 1850  
DB 1462 GGAACCTCGAGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1521

QY 1851 GAGTCTACCAATGAAATACACCAATCTCAGCTCAGCTCAGCTTCCAAACAGAGGACCTGG 1910  
DB 1522 TGAGCTACAGGTAGAAACATCAACACAGATCGTAAAGAGGAGGAGGAGGAGGAGGAG 1581

QY 1911 GCCTTCAGCTCATGAAGAA 1929  
DB 1582 ACGAAAGCAGCAGAGAA 1600

RESULT 5  
US-08-072-070-1  
Sequence 1, Application US/08072070  
Patent No. 5476929  
GENERAL INFORMATION:  
APPLICANT: Briles, David E  
APPLICANT: Yother, Janet L  
APPLICANT: McDaniel, Larry S  
TITLE OF INVENTION: STRUCTURAL GENE OF PNEUMOCOCCAL  
TITLE OF INVENTION: PROTEIN  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Shoemaker and Mattare, Ltd  
STREET: Suite 1203, 2001 Jefferson Davis Highway  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/072,070  
FILING DATE: 19930603  
CLASSIFICATION: 424  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US/07/835,698  
FILING DATE: 12-FEB-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/656,773  
FILING DATE: 15-FEB-1991  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 521-0378  
TELEX: LUKPAT WASHINGTON  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2085 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Streptococcus pneumoniae  
STRAIN: Rxi  
IMMEDIATE SOURCE:  
CLONE: JY2008  
FEATURE:  
NAME/KEY: intron  
LOCATION: 1..2085  
FEATURE:  
NAME/KEY: CDS  
LOCATION: join(127..1983, 1987..1992, 1996..2007, 2011  
LOCATION: ..2025, 2029..2031, 2035..2085)  
US-08-072-070-1

Query Match 1.5%; Score 38.4; DB 1; Length 2085;  
Best Local Similarity 46.3%; Pred. No. 0.23;  
Matches 126; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 1606 GAAACAAAAGCTCTGGAAGATCATCAGAGGATGTGACAAAATATCAGGAAGGATCTCT 1665  
DB 397 GAAAAAGCAGCTCTGGAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 456

QY 1666 GCAGAAAACCCAGTTGAGAACCATATCAATATACACATCAATCAATCAATCAATCAAT 1725  
DB 457 GCCTATCAACAGCTTACAGACAAACCCGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAG 516

QY 1726 CCATTGGATTCCAACTCAGGAGAAAGAAATGACCTCAATCTTGTGCTTGTGGGTT 1785  
DB 517 GCTAAGAACGCGAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 576

QY 1786 CCAGAGAAATCTGCTTCTATCTGAAAAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1845  
DB 577 CCTGAGCCAGAGCAGCTTGGCTGAGACTAGAAAAAATCAGAAAGAGCTAAACAAAAAGCA 636

QY 1846 ACTGAGAGTCTACCAATGAAAATTAACACAA 1877  
DB 637 CCAGAACTTACTAAAAAAGCTAGAAAGCTAA 668

RESULT 6  
US-08-465-746-1  
Sequence 1, Application US/08465746  
Patent No. 5679768  
GENERAL INFORMATION:  
APPLICANT: Briles, David E  
APPLICANT: Yother, Janet L  
APPLICANT: McDaniel, Larry S  
TITLE OF INVENTION: EPITOPIC REGIONS OF PNEUMOCOCCAL SURFACE  
TITLE OF INVENTION: PROTEIN A  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:



```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match	1.5%;	Score 38.4;	DB 1;	Length 2085;
Best Local Similarity	46.3%;	Pred. No. 0.23;		
Matches 126;	Conservative 0;	Mismatches 146;	Indels 0;	Gaps 0;
QY 1606	GAACAAAAGCTCCTGAAGAAATCATCAGAGATGTGCACAAAATATCAGGAAGAGGATATCT	1665		
DB				
QY 397	GA AAAAGCAGGCTCTGAAGAGATGGATAAGCGCATGGCGAGCGATTCACAAAGCTGATCTA	456		
DB				
QY 1666	GCAGAAAACCCAGTTGAGAACCATATCAATATATACAAATCAGATTAAGTTCCACAGCCAAAG	1725		
DB				
QY 457	GCCATATCAACAAGCTCAGACAACCGCGCAAAAGACGACGAGATAAGATGATAGATGAA	516		
DB				
QY 1726	CCATTGGATTCCAACTCAGAGAGAAAGAAATCACCTCAATCTTGATCGCTTTGTGGGGTTT	1785		
DB				
QY 517	GCTAAGAAACGCGAAGAAGAGGCCAAAACATAAATTAATCTGTTCCAGCAATCGTAGTT	576		
DB				
QY 1786	CCAGAAGAAATCGTCTCATCTGTA AAAAGCCAGGAACCGAGAACTTCAGATCAGACATAGC	1845		
DB				
QY 577	CCGTGACCCAGAGAGTTGGCTGGAGACTTAAGAAAAAATCAGAAGAAAGTAAACAAAAAGCA	636		
DB				



```
; LOCATION: ..2025, 2029...2031, 2035...2085)
US-08-214-222-1

Query Match
Best Local Similarity 1.5%; Score 38.4; DB 1; Length 2085;
Matches 126; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 1606 GAAACAAAGCTCCCTGAGAGATCATCAGAGGATGTGACAAAATATCAGGAAGGAGTATCT 1665
DB 397 GAAAGAGCAGCGCTCTGAAGAGATGGATAAGCGAGTGGCAGAGTTCACAAAGCGTATCTA 456

QY 1666 GCAGAAAACCCAGTTGAGAACCATATCAATATAACACATCAGATAAGTTTCACAGCCAAG 1725
DB 457 GCCTATCAACAGCTACAGACAAAGCCGCAAAAGAGCGCAGAGATAAGATGATGATGAA 516

QY 1726 CCATTGGATTCCAACTCAGGAGAAAGAAATGACCTCAATCTTGTGCGCTTTGTGGGTT 1785
DB 517 GCTAAGAAACGGAAGAGGCAAAACTAAATTTAATCTGTTTCGAGCAATGGTAGTT 576

QY 1786 CCAGAAAGATCTGCTTCACTCTGAAGAACCCAAAGGACCAAGAACTTCAGATCAGACTAGC 1845
DB 577 CAGGACCAAGAGCTGCGCTGAGACTAAGAAATAATCAGAAGAGCTTAAACAAAAGCA 636

QY 1846 ACTGAGAGTGTACCAATGAAATAAACACCAA 1877
DB 637 CCAGAACTTACTAAAAACTAGAGAGCTAA 668

RESULT 10
US-08-467-852A-1
; Sequence 1, Application US/08467852A
; Patent No. 5856170
; GENERAL INFORMATION:
; APPLICANT: BRILES, David E.
; APPLICANT: YOTHER, Janet L.
; APPLICANT: MCDANIEL, Larry S.
; TITLE OF INVENTION: STRUCTURAL GENE OF PNEUMOCOCCAL PROTEIN
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FROMMER LAWRENCE & HAUG LLP
; STREET: 745 Fifth Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10151
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,852A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: KOWALSKI, Thomas J.
; REGISTRATION NUMBER: 32,147
; REFERENCE/DOCKET NUMBER: 454312-2064
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-588-0800
; TELEFAX: 212-588-0500
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2085 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; STRAIN: Streptococcus pneumoniae Rxl
; IMMEDIATE SOURCE:

; CLONE: JY2008
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1..2085
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(127..1983, 1987..1992, 1996..2007, 2011
; LOCATION: ..2025, 2029...2031, 2035...2085)
US-08-467-852A-1

Query Match
Best Local Similarity 1.5%; Score 38.4; DB 2; Length 2085;
Matches 126; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 1606 GAAACAAAGCTCCCTGAGAGATCATCAGAGGATGTGACAAAATATCAGGAAGGAGTATCT 1665
DB 397 GAAAGAGCAGCGCTCTGAAGAGATGGATAAGCGAGTGGCAGAGTTCACAAAGCGTATCTA 456

QY 1666 GCAGAAAACCCAGTTGAGAACCATATCAATATAACACATCAGATAAGTTTCACAGCCAAG 1725
DB 457 GCCTATCAACAGCTACAGACAAAGCCGCAAAAGAGCGCAGAGATAAGATGATGATGAA 516

QY 1726 CCATTGGATTCCAACTCAGGAGAAAGAAATGACCTCAATCTTGTGCGCTTTGTGGGTT 1785
DB 517 GCTAAGAAACGGAAGAGGCAAAACTAAATTTAATCTGTTTCGAGCAATGGTAGTT 576

QY 1786 CCAGAAAGATCTGCTTCACTCTGAAGAACCCAAAGGACCAAGAACTTCAGATCAGACTAGC 1845
DB 577 CAGGACCAAGAGCTGCGCTGAGACTAAGAAATAATCAGAAGAGCTTAAACAAAAGCA 636

QY 1846 ACTGAGAGTGTACCAATGAAATAAACACCAA 1877
DB 637 CCAGAACTTACTAAAAACTAGAGAGCTAA 668

RESULT 11
US-08-468-718-1
; Sequence 1, Application US/08468718
; Patent No. 5871943
; GENERAL INFORMATION:
; APPLICANT: Briles, David E
; APPLICANT: Yother, Janet L
; APPLICANT: Mcdaniel, Larry S
; TITLE OF INVENTION: STRUCTURAL GENE OF PNEUMOCOCCAL
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: Suite 1203, 2001 Jefferson Davis Highway
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,718
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/072,068
; FILING DATE: 03 JUNE 1993
; APPLICATION NUMBER: US/07/835,698
; FILING DATE: 12-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/656,773
; FILING DATE: 15-FEB-1991
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 521-0378
```

```

; TELEX: LUKPAT WASHINGTON
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2085 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; STRAIN: Rx1
; IMMEDIATE SOURCE:
; CLONE: JY2008
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1..2085
; NAME/KEY: CDS
; LOCATION: join(127..1983, 1987, 1992, 1996..2007, 2011
; LOCATION: ..2025, 2029..2031, 2035..2085)
; US-08-468-718-1

Query Match 1.5%; Score 38.4; DB 2; Length 2085;
Best Local Similarity 46.3%; Pred. No. 0.23;
Matches 126; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 1606 GAAACAAAGCTCTGGAAGATCATCAGAGGATGTGACAAATATCATCAGGAGGAGTATCT 1665
DB 397 GAAACAAAGCTCTGGAAGATCATCAGAGGATGTGACAAATATCATCAGGAGGAGTATCTA 456
QY 1666 GCAGAAACCCAGTTGAGAACCATATCAATATAACACATCAGATAAGTTTCACAGCCCAAG 1725
DB 457 GCCTATCAACAGCTACAGACAAAGCCGCAAGAGCCGAGATGATGATGAA 516
QY 1726 CCATTGGATTCCAACTCAGGAGAAAGATGACCTCAATCTTGATCGCTCTTTGGGGTT 1785
DB 517 GCTAAGAAACGCGGCTGGAAGAGATGATGATGATGATGATGATGATGATGATGATGAT 576
QY 1786 CCAGAGATCTGCTTCATCTGAAAGCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1845
DB 577 CCTGAGCAGAGCAGTGTGGCTGAGACTAAGAAAAAATCAGAGAAAGCTTAACAAAAAGCA 636
QY 1846 ACTGAGAGTGTACCAATGAAAAATACACCAA 1877
DB 637 CCAGAACTTACTAAAAAAGCTAAGAGAGCTAA 668

RESULT 12
US-08-246-636-1
; Sequence 1, Application US/08246636
; Patent No. 5965141
; GENERAL INFORMATION:
; APPLICANT: Briles, David E.
; APPLICANT: Yother, Janet L.
; APPLICANT: McDaniel, Larry S.
; APPLICANT: Wu, Hong-Yin
; TITLE OF INVENTION: EPTOPIC REGIONS OF PNEUMOCOCCAL SURFACE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: Suite 1203, 2001 Jefferson Davis Highway
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/246, 636
; FILING DATE: 20-MAY-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/656, 773
; FILING DATE: 15-FEB-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/835, 698
; FILING DATE: 12-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/048, 896
; FILING DATE: 20-APR-1993
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; TELEX: LUKPAT WASHINGTON
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2085 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; STRAIN: Rx1
; IMMEDIATE SOURCE:
; CLONE: JY4313
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1..2085
; NAME/KEY: CDS
; LOCATION: join(127..1984)
; US-08-246-636-1

Query Match 1.5%; Score 38.4; DB 2; Length 2085;
Best Local Similarity 46.3%; Pred. No. 0.23;
Matches 126; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 1606 GAAACAAAGCTCTGGAAGATCATCAGAGGATGTGACAAATATCATCAGGAGGAGTATCT 1665
DB 397 GAAACAAAGCTCTGGAAGATCATCAGAGGATGTGACAAATATCATCAGGAGGAGTATCTA 456
QY 1666 GCAGAAACCCAGTTGAGAACCATATCAATATAACACATCAGATAAGTTTCACAGCCCAAG 1725
DB 457 GCCTATCAACAGCTACAGACAAAGCCGCAAGAGCCGAGATGATGATGAA 516
QY 1726 CCATTGGATTCCAACTCAGGAGAAAGATGACCTCAATCTTGATCGCTCTTTGGGGTT 1785
DB 517 GCTAAGAAACGCGGCTGGAAGAGATGATGATGATGATGATGATGATGATGATGATGAT 576
QY 1786 CCAGAGATCTGCTTCATCTGAAAGCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1845
DB 577 CCTGAGCAGAGCAGTGTGGCTGAGACTAAGAAAAAATCAGAGAAAGCTTAACAAAAAGCA 636
QY 1846 ACTGAGAGTGTACCAATGAAAAATACACCAA 1877
DB 637 CCAGAACTTACTAAAAAAGCTAAGAGAGCTAA 668

RESULT 13
US-08-247-491A-1
; Sequence 1, Application US/08247491A
; Patent No. 5965400
; GENERAL INFORMATION:
; APPLICANT: Briles, David E.
; APPLICANT: YOTHER, Janet L.
; TITLE OF INVENTION: STRUCTURAL GENE OF PNEUMOCOCCAL PROTEIN
; NUMBER OF SEQUENCES: 10

```

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: FROMMER LAWRENCE & HAUG LLP
STREET: 745 Fifth Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10151
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/247,491A
FILING DATE: 23-JUN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: KOWALSKI, Thomas J.
REGISTRATION NUMBER: 32,147
REFERENCE/DOCKET NUMBER: 454312-2041
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-588-0800
TELEFAX: 212-588-0500
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2085 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
STRAIN: Streptococcus pneumoniae Rxl
IMMEDIATE SOURCE:
CLONE: JY2008
FEATURE:
NAME/KEY: intron
LOCATION: 1..2085
FEATURE:
NAME/KEY: CDS
LOCATION: join(127..1983, 1987..1992, 1996..2007, 2011
LOCATION: ..2025, 2029..2031, 2035..2085)
US-08-247-491A-1

```

```

Query Match      1.5%; Score 38.4; DB 2; Length 2085;
Best Local Similarity 46.3%; Pred. No. 0.23;
Matches 126; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

```

```

QY 1606 GAAACAAAGCTCCTGAAGATCATCAGAGATGTGACAAATATCAGGAGGAGTATCT 1665
Db      ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
QY 1666 GCAGAAACCCAGTTGAGAACCATATCAATATACAAATCAGATAAGTTTCACAGCCAAAG 1725
Db      || || || || || || || || || || || || || || || || || || || ||
QY 1726 CCATTGATTCCTCACTCAGAGAAAGAAATGACCTCAATCTTGATCCCTCTTGTGGGGTT 1785
Db      || || || || || || || || || || || || || || || || || || || ||
QY 517 GCTAAGAAACGCGAAGAGAGGCAAACTAAATTTAATCTGTTTCGAGCAATGTTAGTT 576
Db      || || || || || || || || || || || || || || || || || || || ||
QY 1786 CCAGAGAAATCTGCTTCATCTGAAAGCCAGCAAGCAAACTTCAGATCAGACTAGC 1845
Db      || || || || || || || || || || || || || || || || || || || ||
QY 577 CTTGAGCCAGAGCAGTTGGCTGAGACTAAGAAATATCAGAAAGCTAAACAAAGCA 636
QY 1846 ACTGAGAGTCTTACCAATGAAATACACCA 1877
Db      || || || || || || || || || || || || || || || || || || || ||
QY 637 CCAGAACTTACTAAAAAACTAGAGAAGCTAA 668
Db      || || || || || || || || || || || || || || || || || || || ||

```

RESULT 14

US-08-319-795-1

; Sequence 1, Application US/08319795

```

; Patent No. 5980909
; GENERAL INFORMATION:
; APPLICANT: Briles, David E.
; APPLICANT: Yoher, Janet L.
; APPLICANT: McDaniel, Larry S
; TITLE OF INVENTION: Epitopic Regions of Pneumococcal Surface
; TITLE OF INVENTION: Protein A
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheomaker and Mattare, Ltd.
; STREET: 1203 Crystal Plaza Bldg. 1, 2001 Jefferson
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/319,795
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/246,636
; FILING DATE: 20-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/048,896
; FILING DATE: 20-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/835,698
; FILING DATE: 12-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/656,773
; FILING DATE: 15-FEB-1991
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2085 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; STRAIN: Rxl
; IMMEDIATE SOURCE:
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1..2085
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(127..1984)
; US-08-319-795-1

```

```

Query Match      1.5%; Score 38.4; DB 2; Length 2085;
Best Local Similarity 46.3%; Pred. No. 0.23;
Matches 126; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

```

```

QY 1606 GAAACAAAGCTCCTGAAGATCATCAGAGATGTGACAAATATCAGGAGGAGTATCT 1665
Db      ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
QY 397 GAAAGACGCGTCTGAAGAGATGGATAGCGAGTGGCAGCGTTCACAAAGCGTATCTA 456
Db      || || || || || || || || || || || || || || || || || || || ||
QY 1666 GCAGAAACCCAGTTGAGAACCATATCAATATACAAATCAGATAAGTTTCACAGCCAAAG 1725
Db      || || || || || || || || || || || || || || || || || || || ||
QY 457 GCCTATCAACAGCTACAGACAAAGCCGCAAGCGCAAGATGATAGTAA 516
Db      || || || || || || || || || || || || || || || || || || || ||

```





Db 511 TGAGGCTACAGGTTAGAAAAAATCAAGACAGATCGTAAAAAAGCAGAGAAGAGCTAA 570  
QY 1911 GCCTTCAGCTCATGAAGAA 1929  
Db 571 ACAGAAAGCAGCAGAGAA 589

RESULT 19  
US-08-961-527-134  
; Sequence 134, Application US/08961527  
; Patent No. 6420135  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunsch  
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 391  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION NUMBER: US/08/961,527  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brookes, A. Anders  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB340P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 134:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 12665 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-08-961-527-134

Query Match 1.4%; Score 37.4; DB 4; Length 12665;  
Best Local Similarity 44.8%; Pred. No. 1.5;  
Matches 143; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

QY 1611 AAAAGCTCTCAGAGTATCATCAGAGGATGTGCAAAATATCAGGAGGAGTATCGCAGA 1670  
Db 1560 AAAAAGTAGCAGAGCTGAGAGAAGGTTGAAGAGCTAGAAAAGCGGAGATCA 1619

QY 1671 AAACCCAGTTGAGACCATATCAATATACACAAATCAGATAGGTTACAGCCAGCCATT 1730  
Db 1620 AAAAGAAGAAGATCGCGTAACTACCCCAACCAATACTTACAAAAGCGTTGAACTTGAAT 1679

QY 1731 GGATTCACCTCAGAGAGAAGATGACCTCAATCTTATCGCTCTTGGGGTTCAGA 1790  
Db 1680 TGCTGAGTCCGATGTGGAAGTTAAAAAAGCGGAGCTTGAACCTAGTAAAAAGAGAGCTAA 1739

QY 1791 AGAATCTGCTTCATCTGAAAAAGCCAGGAACCAACTTCAGATCAGCTAGCACTGA 1850  
Db 1740 GGAACCTGAAACGAGGAAAAGATTAAACAGCAAAAGCGGAGTTGAGATAAAAAGC 1799

QY 1851 GAGTCTACCAATGAAAATAACACCAATCTTGAGCCTCAGTTCCAAACAGAGCCACTGG 1910  
Db 1800 TGAGGCTACAGGTTAGAAAAAATCAAGACAGATCGTAAAAAAGCAGAGAAGAGCTAA 1859

QY 1911 GCCTTCAGCTCATGAAGAA 1929  
Db 1860 ACAGAAAGCAGCAGAGAA 1878

RESULT 20  
US-09-007-005-17  
; Sequence 17, Application US/09007005B  
; Patent No. 6258558  
; GENERAL INFORMATION:  
; APPLICANT: Szostak, Jack W.  
; APPLICANT: Roberts, Richard W.  
; APPLICANT: Liu, Rihe  
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN  
; FILE REFERENCE: 00786/350003  
; CURRENT APPLICATION NUMBER: US/09/007,005B  
; CURRENT FILING DATE: 1998-01-14  
; EARLIER APPLICATION NUMBER: 60/035,963  
; EARLIER FILING DATE: 1997-01-27  
; EARLIER APPLICATION NUMBER: 60/064,491  
; EARLIER FILING DATE: 1997-11-06  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 17  
; LENGTH: 289  
; TYPE: RNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Translation template  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(289)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-007-005-17

Query Match 1.4%; Score 35.8; DB 4; Length 289;  
Best Local Similarity 5.5%; Pred. No. 0.39;  
Matches 13; Conservative 101; Mismatches 121; Indels 0; Gaps 0;

QY 911 AACCCAGTAAAGCGTTTGAGACTTCCTGGTGAATGTCAGATACCTGACCCAGAGCAA 970  
Db 8 ARCARARURURARURURARURARURARURARURARURARURARURARURARURAR 67

QY 971 GCCCGGAGTGACGAGAAAGAGATGGAGAGCAGATGCCCAATGTGTCATTTGTCAGA 1030  
Db 68 SRNR 127

QY 1031 GAATGCTGATATGTTATCAAGATGTTTGAAGAAGCAAGTGAGGTGCACAAAGCAATA 1090  
Db 128 SRNR 187

QY 1091 GAGGACGAGGAGATCTCGACCCAGAGGTGGAAAGTCAATCAGATATTTCAAC 1145  
Db 188 SRNR 242

RESULT 21  
US-09-244-796-17  
; Sequence 17, Application US/09244796  
; Patent No. 6281344  
; GENERAL INFORMATION:  
; APPLICANT: Szostak, Jack W.  
; APPLICANT: Roberts, Richard W.  
; APPLICANT: Liu, Rihe  
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN  
; FILE REFERENCE: 00786/350007  
; CURRENT APPLICATION NUMBER: US/09/244,796  
; CURRENT FILING DATE: 1999-02-05  
; EARLIER APPLICATION NUMBER: 60/035,963  
; EARLIER FILING DATE: 1997-01-27  
; EARLIER APPLICATION NUMBER: 60/064,491



```

RESULT 22
US-09-308-022-5
; Sequence 5, Application US/09308022
; Patent No. 6291654
;
; GENERAL INFORMATION:
;
; APPLICANT: REGENTS OF THE UNIVERSITY OF MINNESOTA, ET AL.
; TITLE OF INVENTION: C3 BINDING PROTEIN OF STREPTOCOCCUS
; TITLE OF INVENTION: PNEUMONIAE
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
;
; ADDRESSEE: MUEITING, RAASCH & GEBHARDT, P.A.
; STREET: 119 NO. 6291654th Fourth Street, Suite 203
; CITY: Minneapolis
; STATE: Minnesota
; COUNTRY: USA
;
; ZIP: 55401
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/308,022
; FILING DATE:
;
; CLASSIFICATION:
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/20586
; FILING DATE: 12-NOV-1997
; APPLICATION NUMBER: 60/029,444
; FILING DATE: 12-NOV-1996
; APPLICATION NUMBER: 60/038,086
; FILING DATE: 18-FEB-1997
; APPLICATION NUMBER: 60/059,368
; FILING DATE: 19-SEP-1997
; APPLICATION NUMBER: 60/062,473

```

Matches	134;	Conservative	0;	Mismatches	164;	Indels	0;	Gaps	0;
QY	1611	AAAAGCTCCTGAGAAATCATCAGAGATGTGACAAAATATCATGGAAGGAGTATCTGCAGA	1670						
Db	300	AAAAAGGTAGCAGAGCTGAGAGAAGGTTGAAGAGCTGAGAAAAAGCCAAAGGATCA	359						
QY	1671	AAACCCAGTTGAGAACCATATCAATATAACACAAATCAGATAAGTTTCACACGCCAGCCATT	1730						
Db	360	AAAAAGAAGATCGCGTAACCTACCCACCATACTTACAAAACGGTTGACCTTGAAT	419						
QY	1731	GGATTCCAACTCAGGAGAAAGAANAATGACCTCAATCTTGATCGCTCTTGTCGGGTTCAGA	1790						
Db	420	TGCTGAGTCGGATGTGAAAGTTTAAAGAAGCGGAGCTTGAACTAGTAAAGAGGAAGCTAA	479						
QY	1791	AGAATCTGCTTCTGAAAGGCCAAGGAACACAGAAACTTCAGATCATGACTAGCCTGA	1850						
Db	480	GGAACCTCGACGAGGAGAAAAAATTAAGCAAGCAAAAGCGAAAGTTGAGAGTAAAAAGC	539						
QY	1851	GAGTGCTACCAATGAAATTAACACCAATCTTGAGCTCAGTTCCTCCAAACAGAAGCCACT	1908						
Db	540	TGAGGCTACAGGTTAGAAAACATCAAGACAGATCTGTAATAAAGCAGAGAAGAGCT	597						

```

RESULT 23
US-09-385-982-267
; Sequence 267, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; TITLE OF INVENTION: PRODUCTS: II
; FILE REFERENCE: CCDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 267
; LENGTH: 598
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(598)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-267

```

Query Match	1.4%	Score 35.4;	DB 4;	Length 598;
Best Local Similarity	50.9%	Pred. No. 0.83;		



RESULT 26  
US-08-299-074A-1/c  
; Sequence 1, Application US/08299074A  
; Patent No. 5955263  
; GENERAL INFORMATION:  
; APPLICANT: Vogelstein, Bert  
; APPLICANT: Kinzler, Michael  
; APPLICANT: Sherman, Michael  
; TITLE OF INVENTION: SEQUENCE SPECIFIC DNA BINDING  
; NUMBER OF SEQUENCES: 41  
; ADDRESS: Banner & Witcoff  
; STREET: 1001 G Street, NW  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20001  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/299,074A  
; FILING DATE: 01-SEP-1994  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/860,758  
; FILING DATE: 31-MAR-1992  
; APPLICATION NUMBER: 07/715,182  
; FILING DATE: 14-JUN-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kagan, Sarah A  
; REGISTRATION NUMBER: 32141  
; REFERENCE/DOCKET NUMBER: 01107.47071  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-508-9100  
; TELEFAX: 202-508-9299  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 405 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-299-074A-1

Query Match 1.3%; Score 34.6; DB 2; Length 405;  
Best Local Similarity 49.7%; Pred. No. 1.1;  
Matches 88; Conservative 0; Mismatches 89; Indels 0; Gaps 0;  
QY 1588 CCACCAGAGGAGACATGTAACAAAGCTCTCTGAAAGATCATCAGAGGATGTGACAAA 1647  
Db 263 CAACAAGAGTGAACCTGTCTCAAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 204  
QY 1648 TATCAGAGAGGAGTATCTGAGAAACCCAGTTGAGAACCATATCATATATACACAATCA 1707  
Db 203 GAAGAAG 144  
QY 1708 GATAAGTTCACAGCCAGCATTTGGATTCCAACTCAGGAGAGAGAGAGAGAGAGAGAG 1764  
Db 143 AGAAG 87

RESULT 27  
US-09-399-773-1/c  
; Sequence 1, Application US/09399773  
; Patent No. 6245515  
; GENERAL INFORMATION:  
; APPLICANT: Vogelstein, Bert

; APPLICANT: Kinzler, Kenneth  
; APPLICANT: Sherman, Michael  
; TITLE OF INVENTION: SEQUENCE SPECIFIC DNA BINDING  
; NUMBER OF SEQUENCES: 41  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Banner & Witcoff  
; STREET: 1001 G Street, NW  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20001  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/399,773  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/299,074  
; FILING DATE:  
; APPLICATION NUMBER: 07/715,182  
; FILING DATE: 14-JUN-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kagan, Sarah A  
; REGISTRATION NUMBER: 32141  
; REFERENCE/DOCKET NUMBER: 01107.47071  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-508-9100  
; TELEFAX: 202-508-9299  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 405 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-09-399-773-1

Query Match 1.3%; Score 34.6; DB 4; Length 405;  
Best Local Similarity 49.7%; Pred. No. 1.1;  
Matches 88; Conservative 0; Mismatches 89; Indels 0; Gaps 0;  
QY 1588 CCACCAGAGGAGACATGTAACAAAGCTCTCTGAAAGATCATCAGAGGATGTGACAAA 1647  
Db 263 CAACAAGAGTGAACCTGTCTCAAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 204  
QY 1648 TATCAGAGAGGAGTATCTGAGAAACCCAGTTGAGAACCATATCATATATACACAATCA 1707  
Db 203 GAAGAAG 144  
QY 1708 GATAAGTTCACAGCCAGCATTTGGATTCCAACTCAGGAGAGAGAGAGAGAGAGAGAG 1764  
Db 143 AGAAG 87

RESULT 28  
US-08-606-288-6/c  
; Sequence 6, Application US/08606288  
; Patent No. 5955087  
; GENERAL INFORMATION:  
; APPLICANT: Whittle, N.R.  
; APPLICANT: Carmichael, J.P.  
; APPLICANT: Connor, S.E.  
; APPLICANT: Thompson, H.S.G.  
; APPLICANT: Wilson, M.J.  
; TITLE OF INVENTION: Polypeptides useful as immunotherapeutic  
; TITLE OF INVENTION: Agents, and Methods of Polypeptide Preparation  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
STREET: Suite 3400, Four Embarcadero Center  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/606,288  
FILING DATE: 23-FEB-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9503786.7  
FILING DATE: 24-FEB-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/000034  
FILING DATE: 08-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9515478.7  
FILING DATE: 28-JUL-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Walter H. Dreger  
REGISTRATION NUMBER: 24,190  
REFERENCE/DOCKET NUMBER: A-63284/WH  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1815 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 7..1812  
US-08-606-288-6

Query Match 1.3%; Score 34.6; DB 2; Length 1815;  
Best Local Similarity 50.3%; Pred. No. 3;  
Matches 85; Conservative 0; Mismatches 84; Indels 0; Gaps 0;  
QY 1588 CCACGAGGAGGAGCAGTGAACAAAGCTCTGAGGAATCATCAGAGGATGTGACAAA 1647  
DB 350 CCACGTAGTAATAGAGGTTTGCAGAGTTCTTAAGGGAACATAGCCAGTACGCCCA 291  
QY 1648 TATCAGGAGGAGTATCTGCAGAAACCCAGTTGAGAACCATATCAATATACAAATCA 1707  
DB 290 GTGCGGAACCGGTGCCTATACCAACCTCGGAAGACACCCCAACTCCCCATTTT 231  
QY 1708 GATAGTTTCAGCAACCCATTTGATTCCTCACTCAGGAGGAAGATG 1756  
DB 230 AATATTGATCTGCAATGGTGTGTGTTCCACCTTAGGAATTACATCTG 182

RESULT 29  
US-09-347-483-6/C  
; Sequence 6, Application US/09347483  
; Patent No. 6123948  
; GENERAL INFORMATION:  
; APPLICANT: Whittle, N.R.  
; APPLICANT: Carmichael, J.P.  
; APPLICANT: Connor, S.E.  
; APPLICANT: Thompson, H.S.G.  
; APPLICANT: Wilson, W.J.  
; TITLE OF INVENTION: Polypeptides Useful as Immunotherapeutic  
; TITLE OF INVENTION: Agents, and Methods of Polypeptide Preparation  
; NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
STREET: Suite 3400, Four Embarcadero Center  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/347,483  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/606,288  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/000034  
FILING DATE: 08-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9515478.7  
FILING DATE: 28-JUL-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Walter H. Dreger  
REGISTRATION NUMBER: 24,190  
REFERENCE/DOCKET NUMBER: A-63284/WH  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1815 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 7..1812  
US-09-347-483-6

Query Match 1.3%; Score 34.6; DB 3; Length 1815;  
Best Local Similarity 50.3%; Pred. No. 3;  
Matches 85; Conservative 0; Mismatches 84; Indels 0; Gaps 0;  
QY 1588 CCACGAGGAGGAGCAGTGAACAAAGCTCTGAGGAATCATCAGAGGATGTGACAAA 1647  
DB 350 CCACGTAGTAATAGAGGTTTGCAGAGTTCTTAAGGGAACATAGCCAGTACGCCCA 291  
QY 1648 TATCAGGAGGAGTATCTGCAGAAACCCAGTTGAGAACCATATCAATATACAAATCA 1707  
DB 290 GTGCGGAACCGGTGCCTATACCAACCTCGGAAGACACCCCAACTCCCCATTTT 231  
QY 1708 GATAGTTTCAGCAACCCATTTGATTCCTCACTCAGGAGGAAGATG 1756  
DB 230 AATATTGATCTGCAATGGTGTGTGTTCCACCTTAGGAATTACATCTG 182

RESULT 30  
US-09-007-005-32  
; Sequence 32, Application US/09007005B  
; Patent No. 6258558  
; GENERAL INFORMATION:  
; APPLICANT: Szostak, Jack W.  
; APPLICANT: Roberts, Richard W.  
; APPLICANT: Liu, Rihe  
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN  
; TITLE OF INVENTION: FUSIONS  
; FILE REFERENCE: 00786/350003  
; CURRENT APPLICATION NUMBER: US/09/007,005B

; CURRENT FILING DATE: 1998-01-14  
; EARLIER APPLICATION NUMBER: 60/035,963  
; EARLIER FILING DATE: 1997-01-27  
; EARLIER APPLICATION NUMBER: 60/064,491  
; EARLIER FILING DATE: 1997-11-06  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 32  
; LENGTH: 248  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-09-007-005-32

Query Match 1.3%; Score 34.4; DB 4; Length 248;  
Best Local Similarity 20.9%; Pred. No. 0.95;  
Matches 36; Conservative 64; Mismatches 72; Indels 0; Gaps 0;  
QY 847 AAAGATGATACACGACGAGAACTTAAACTCTTCTCGGAGAGAGAGAGAGAGTTG 906  
Db 67 RARGRARCRU 126  
QY 907 GCACAAACACGAGTTAAGCGTTTGAAGACTTCGTTGTTGATTCAGATCTGGACCCAGA 966  
Db 127 RRCRURGRURARCR 186  
QY 967 GCAAGCGCGGAGAGTGAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1018  
Db 187 RCRARCRU 238

RESULT 31  
US-09-244-796-32  
; Sequence 32, Application US/09244796  
; Patent No. 6281344  
; GENERAL INFORMATION:  
; APPLICANT: Szostak, Jack W.  
; APPLICANT: Roberts, Richard W.  
; APPLICANT: Liu, Rihe  
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN  
; TITLE OF INVENTION: FUSIONS  
; FILE REFERENCE: 00786/350007  
; CURRENT APPLICATION NUMBER: US/09/244,796  
; CURRENT FILING DATE: 1999-02-05  
; EARLIER APPLICATION NUMBER: 60/035,963  
; EARLIER FILING DATE: 1997-01-27  
; EARLIER APPLICATION NUMBER: 60/064,491  
; EARLIER FILING DATE: 1997-11-06  
; EARLIER FILING DATE: 1998-01-14  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 32  
; LENGTH: 248  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-09-244-796-32

Query Match 1.3%; Score 34.4; DB 4; Length 248;  
Best Local Similarity 20.9%; Pred. No. 0.95;  
Matches 36; Conservative 64; Mismatches 72; Indels 0; Gaps 0;  
QY 847 AAAGATGATACACGACGAGAACTTAAACTCTTCTCGGAGAGAGAGAGAGAGTTG 906  
Db 67 RARGRARCRU 126  
QY 907 GCACAAACACGAGTTAAGCGTTTGAAGACTTCGTTGTTGATTCAGATCTGGACCCAGA 966  
Db 127 RRCRURGRURARCR 186  
QY 967 GCAAGCGCGGAGAGTGAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1018  
Db 187 RCRARCRU 238

RESULT 32  
US-09-007-005-3  
; Sequence 3, Application US/09007005B  
; Patent No. 6258558  
; GENERAL INFORMATION:  
; APPLICANT: Szostak, Jack W.  
; APPLICANT: Roberts, Richard W.  
; APPLICANT: Liu, Rihe  
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN  
; TITLE OF INVENTION: FUSIONS  
; FILE REFERENCE: 00786/350003  
; CURRENT APPLICATION NUMBER: US/09/007,005B  
; CURRENT FILING DATE: 1998-01-14  
; EARLIER APPLICATION NUMBER: 60/035,963  
; EARLIER FILING DATE: 1997-01-27  
; EARLIER APPLICATION NUMBER: 60/064,491  
; EARLIER FILING DATE: 1997-11-06  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 277  
; TYPE: RNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Translation template  
US-09-007-005-3

Query Match 1.3%; Score 34.4; DB 4; Length 277;  
Best Local Similarity 20.9%; Pred. No. 1;  
Matches 36; Conservative 64; Mismatches 72; Indels 0; Gaps 0;  
QY 847 AAAGATGATACACGACGAGAACTTAAACTCTTCTCGGAGAGAGAGAGAGAGTTG 906  
Db 67 RARGRARCRU 126  
QY 907 GCACAAACACGAGTTAAGCGTTTGAAGACTTCGTTGTTGATTCAGATCTGGACCCAGA 966  
Db 127 RRCRURGRURARCR 186  
QY 967 GCAAGCGCGGAGAGTGAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1018  
Db 187 RCRARCRU 238

RESULT 33  
US-09-244-796-3  
; Sequence 3, Application US/09244796  
; Patent No. 6281344  
; GENERAL INFORMATION:  
; APPLICANT: Szostak, Jack W.  
; APPLICANT: Roberts, Richard W.  
; APPLICANT: Liu, Rihe  
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN  
; TITLE OF INVENTION: FUSIONS  
; FILE REFERENCE: 00786/350007  
; CURRENT APPLICATION NUMBER: US/09/244,796  
; CURRENT FILING DATE: 1999-02-05  
; EARLIER APPLICATION NUMBER: 60/035,963  
; EARLIER FILING DATE: 1997-01-27  
; EARLIER APPLICATION NUMBER: 60/064,491  
; EARLIER FILING DATE: 1997-11-06  
; EARLIER APPLICATION NUMBER: 09/007,005  
; EARLIER FILING DATE: 1998-01-14  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 277  
; TYPE: RNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Translation template  
US-09-244-796-3



```

; TISSUE TYPE: fiber cells
; IMMEDIATE SOURCE:
; LIBRARY: CKFB15A1
; CLONE: H6
US-07-885-970A-3

```

Query Match 1.3%; Score 34.4; DB 1; Length 913;  
Best Local Similarity 46.9%; Pred. No. 2.2;  
Matches 107; Conservative 0; Mismatches 121; Indels

	1105	TCTCGACCCAGAGTGGAACAAAGTCAAATCAGATTATTTCAACTCTTCTACGGTCCCACATCA	1164
QY			
	297	TCTCCACCCCCAGTTTTCAACTCTCCACCAGTTCTCTGCTGTGGAATCCACCACCT	356
Db			
	1165	AGTCCTGATTTGGAGTGAGTGAACCTGCAATGGAAGTAGATACTCCAGCTGGAACAATTT	1224
QY			
	357	GCTTCTCCTCCTCTGGAATCTCACCTCCAGCTTCTCCACCTCTGTCATCTCTCCACCA	416
Db			
	1225	CATTACGCCCTTCTACATCCTCTACAATGTCAAGCTCAGGCTCATTCGCATCATCTCCACACA	1284
QY			
	417	GCTTCTCCACCTCCGCCACTCCACCACCTGCAACCCACCGCCAGCAACTCTCTCTCTCT	476
Db			
	1285	GAAGCGCTCATTTACTCCTTTGGTCTATCTTCTCCAGACAGTGAACAA	1332
QY			
	477	GCTACCCACACACAGCTCCGATTTGGTTTCTCTCCAGCCACAGTCCCA	524
Db			

RESULT 36

US-08-298-687A-3  
; Sequence 3, Application US/08298687A  
; Patent No. 5521078

```

: GENERAL INFORMATION:
:
: APPLICANT: John, Mallyakal E.
:
: TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON
:
: TITLE OF INVENTION: PLANTS FOR ALTERED FIBER
:
: NUMBER OF SEQUENCES: 33
:
: CORRESPONDENCE ADDRESS:
:
: ADDRESSEE: Nicholas J. Seay, Quarles & Brady
:
: STREET: P. O. Box 2113, First Wisconsin Plaza
:
: CITY: Madison
:
: STATE: Wisconsin
:
: COUNTRY: USA
:
: ZIP: 53701

```

```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/298,687A
;
```

```

;
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/617,239
; FILING DATE: 21-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/253,243
; FILING DATE: 04-OCT-1988
;

```

/ ZERBING DATA: 04 OCT 1988  
 / ATTORNEY/AGENT INFORMATION:  
 / NAME: Seay, Nicholas J.  
 / REGISTRATION NUMBER: 27,386  
 / TELECOMMUNICATION INFORMATION:  
 / TELEPHONE: (608) 283-2478  
 / TELEFAX: (608) 251-5139  
 / INFORMATION FOR SEQ ID NO: 3:

```

SEQUENCE CHARACTERISTICS:
LENGTH: 913 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO

```

## ORIGINAL SOURCE:

ORGANISM: Cossypium hirsutum  
STRAIN: Coker 312  
DEVELOPMENTAL STAGE: 15 day old fiber cells  
TISSUE TYPE: fiber cells  
IMMEDIATE SOURCE:  
LIBRARY: CKFB151A  
CLONE: H6  
US-08-298-687A-3

```
Query Match      1.3%; Score 34.4; DB 1; Length 913;
Best Local Similarity 46.9%; Pred. No. 2.2;
Matches 107; Conservative 0; Mismatches 121; Indels 0; Gaps 0;
```

[illegible]

RESULT 37

US-08-530-797-2  
; Sequence 2, Application US/08530797  
; Patent No. 5597718  
; GENERAL INFORMATION:  
; APPLICANT: John, Maliyakal E.  
; APPLICANT: Umbeck, Paul F.  
; APPLICANT: Brill, Winston J.  
; TITLE OF INVENTION: GENETICALLY ENGINEERED COTTON PLANTS  
; TITLE OF INVENTION: FOR ALTERED FIBER  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Quarles and Brady  
; STREET: P.O BOX 2113  
; STREET: FIRST WISCONSIN PLAZA  
; CITY: MADISON  
; STATE: WISCONSIN  
; COUNTRY: U.S.A.  
; 23D. 53701

```

? ZIP: 53701
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette - 3.50 inch, 800Kb storage
? COMPUTER: Apple Macintosh
? OPERATING SYSTEM: Macintosh
? SOFTWARE: Microsoft Word 4.0
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/530,797
? FILING DATE: 20-SEP-1995
?
```

CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/617,239  
FILING DATE: 21-NOV-90  
APPLICATION NUMBER: US 07/253,243  
FILING DATE: 04-OCT-88  
ATTORNEY/AGENT INFORMATION:

```

; NAME: Nicholas J. Seay
; REGISTRATION NUMBER: 27,386
; REFERENCE/DOCKET NUMBER: 1122990245
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 913 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
;

```





```

:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Microsoft word
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/885,970A
: FILING DATE: 19920518
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/617,239
: FILING DATE: 21-NOV-1990
:

```

Search completed: March 26, 2003, 02:00:06  
Job time : 173 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 26, 2003, 02:00:09 ; Search time 315 Seconds  
(without alignments)  
6969.874 Million cell updates/sec

Title: US-09-781-693A-1

Perfect score: 2580

Sequence: 1 atctctcggggtggctctca.....atgaaatgaggatgaggaa 2580

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 574371 seqs, 425486471 residues

Total number of hits satisfying chosen parameters: 1148742

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 120 summaries

Database :

Published Applications NA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/PTCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/PTCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2580	100.0	2580	9	US-09-781-693A-1
2	2580	100.0	3016	9	US-09-781-693A-3
3	1542.4	59.8	1942	9	US-09-922-846-200
4	328	12.7	332	10	US-09-864-761-19286
5	261	10.1	261	10	US-09-864-761-17295
6	255.8	9.9	481	10	US-09-864-761-2560
7	241.8	9.4	409	10	US-09-983-965-58
8	186	7.2	186	10	US-09-864-761-17296
9	160	6.2	467	10	US-09-864-761-499
10	141	5.5	468	10	US-09-864-761-500
11	71.2	2.8	356	10	US-09-770-791-605
12	59	2.3	257	10	US-09-983-965-46
13	54.6	2.1	2679	10	US-09-925-300-645
14	41	1.6	2276	10	US-09-828-447-6
15	39.2	1.5	640661	10	US-09-790-988-1
16	38.6	1.5	1945	10	US-09-925-300-145
17	37.4	1.4	1360	10	US-09-765-272-37
18	37.4	1.4	2794	10	US-09-917-800A-1663
19	36.6	1.4	574	9	US-10-184-644-234

36.6	1.4	671	9	US-10-184-644-346	Sequence 346, App
36.2	1.4	276	10	US-09-864-761-20595	Sequence 20595, A
36.2	1.4	295	10	US-09-923-876-4276	Sequence 4276, Ap
36.2	1.4	462	10	US-09-864-761-3829	Sequence 3829, Ap
35.6	1.4	561	10	US-09-998-598-1297	Sequence 1297, Ap
35.4	1.4	528	9	US-10-152-661-618	Sequence 618, App
35.4	1.4	594	9	US-09-866-050A-618	Sequence 618, App
35.4	1.4	594	10	US-09-925-301-690	Sequence 690, App
35.4	1.4	895	10	US-09-770-445-480	Sequence 480, App
35.4	1.4	1056	12	US-10-044-090-8	Sequence 8, Appli
35.4	1.4	1188	10	US-09-998-598-368	Sequence 368, App
35.4	1.4	1548	10	US-09-998-598-357	Sequence 357, App
35.2	1.4	3972	10	US-09-801-368-33	Sequence 33, Appl
35	1.4	429	10	US-09-860-432-73	Sequence 73, Appl
35	1.4	453	10	US-09-998-598-61	Sequence 61, Appl
35	1.4	559	10	US-09-864-761-7684	Sequence 7684, Ap
35	1.4	4301	9	US-09-989-919-69	Sequence 69, Appl
35	1.4	6823	9	US-09-989-920-16	Sequence 16, Appl
34.8	1.3	929	10	US-09-867-550-1267	Sequence 1267, Ap
34.8	1.3	3953	10	US-09-864-761-19041	Sequence 19041, A
34.8	1.3	7287	10	US-09-070-927A-210	Sequence 210, App
34.6	1.3	405	9	US-09-813-824A-1	Sequence 1, Appli
34.6	1.3	2000	9	US-09-938-842A-4180	Sequence 4180, Ap
34.6	1.3	11802	7	US-08-781-986A-70	Sequence 70, Appl
34.4	1.3	309	10	US-09-764-887-368	Sequence 368, App
34.4	1.3	309	10	US-09-764-887-369	Sequence 369, App
34.4	1.3	309	10	US-09-764-887-370	Sequence 370, App
34.4	1.3	589	10	US-09-864-761-6977	Sequence 6977, App
34.4	1.3	701	10	US-09-864-761-19757	Sequence 19757, A
34.4	1.3	701	10	US-09-864-761-26789	Sequence 26789, A
34.4	1.3	963	10	US-09-864-761-10148	Sequence 10148, A
34.4	1.3	1336	10	US-09-799-777-114	Sequence 114, App
34.4	1.3	1341	10	US-09-814-122-14	Sequence 14, Appl
34.4	1.3	1646	10	US-09-816-828-3	Sequence 3, Appli
34.4	1.3	1934	10	US-09-864-761-2976	Sequence 2976, Ap
34.4	1.3	64667	9	US-10-274-409-3	Sequence 3, Appli
34.2	1.3	848	10	US-09-770-445-632	Sequence 632, App
34.2	1.3	4437	10	US-09-801-368-51	Sequence 51, Appl
34	1.3	320	10	US-09-764-887-40	Sequence 40, Appl
34	1.3	503	9	US-09-981-876-19	Sequence 19, Appl
34	1.3	503	9	US-09-148-545-19	Sequence 19, Appl
33.8	1.3	420	10	US-09-864-761-5017	Sequence 5017, Ap
33.8	1.3	452	10	US-09-864-761-10129	Sequence 10129, A
33.8	1.3	458	9	US-10-184-644-546	Sequence 546, App
33.8	1.3	811	9	US-10-184-644-414	Sequence 414, App
33.8	1.3	13321	7	US-08-781-986A-4	Sequence 4, Appli
33.6	1.3	189	9	US-09-864-761-23708	Sequence 23708, A
33.6	1.3	499	10	US-10-040-739-194	Sequence 194, App
33.6	1.3	526	10	US-09-864-761-7895	Sequence 7895, Ap
33.6	1.3	2000	9	US-09-938-842A-3975	Sequence 3975, Ap
33.6	1.3	4656	10	US-09-864-761-19341	Sequence 19341, A
33.6	1.3	6233	12	US-10-044-090-371	Sequence 371, App
33.6	1.3	7100	10	US-09-932-183A-1	Sequence 1, Appli
33.6	1.3	465327	10	US-09-933-267A-1	Sequence 1, Appli
33.4	1.3	497	10	US-09-783-590-8295	Sequence 8295, Ap
33.4	1.3	734	9	US-10-184-644-458	Sequence 458, App
33.4	1.3	3458	7	US-08-781-986A-317	Sequence 317, App
33.2	1.3	282	10	US-09-922-261-205	Sequence 205, App
33.2	1.3	306	10	US-09-922-261-203	Sequence 203, App
33.2	1.3	696	10	US-09-922-261-193	Sequence 193, App
33.2	1.3	699	10	US-09-922-261-191	Sequence 191, App
33.2	1.3	707	9	US-10-040-739-1385	Sequence 1385, Ap
33.2	1.3	710	10	US-09-922-261-189	Sequence 189, App
33.2	1.3	774	10	US-09-922-261-187	Sequence 187, App
33.2	1.3	819	10	US-09-922-261-185	Sequence 185, App
33.2	1.3	957	10	US-09-735-705-113	Sequence 113, App
33.2	1.3	957	10	US-09-850-710A-113	Sequence 113, App
33.2	1.3	957	10	US-09-897-778-113	Sequence 113, App
33.2	1.3	1669	10	US-09-922-261-184	Sequence 184, App
33.2	1.3	1915	9	US-09-938-842A-3951	Sequence 3951, Ap
33.2	1.3	2870	9	US-10-174-590-473	Sequence 473, App
33.2	1.3	2870	9	US-10-176-758-473	Sequence 473, App
33.2	1.3	2870	9	US-10-175-737-473	Sequence 473, App

c	93	33.2	1.3	2870	9	US-10-173-706-473	Sequence 473, App
c	94	33.2	1.3	2870	9	US-10-175-738-473	Sequence 473, App
c	95	33.2	1.3	2870	9	US-10-175-752-473	Sequence 473, App
c	96	33.2	1.3	2870	9	US-10-176-482-473	Sequence 473, App
c	97	33.2	1.3	2870	9	US-10-176-757-473	Sequence 473, App
c	98	33.2	1.3	2870	9	US-10-176-913-473	Sequence 473, App
c	99	33.2	1.3	2870	9	US-10-180-552-473	Sequence 473, App
c	100	33.2	1.3	2870	9	US-10-180-557-473	Sequence 473, App
c	101	33.2	1.3	2870	9	US-10-173-700-473	Sequence 473, App
c	102	33.2	1.3	2870	9	US-10-174-572-473	Sequence 473, App
c	103	33.2	1.3	2870	9	US-10-174-579-473	Sequence 473, App
c	104	33.2	1.3	2870	9	US-10-174-582-473	Sequence 473, App
c	105	33.2	1.3	2870	9	US-10-174-588-473	Sequence 473, App
c	106	33.2	1.3	2870	9	US-10-175-739-473	Sequence 473, App
c	107	33.2	1.3	2870	9	US-10-175-740-473	Sequence 473, App
c	108	33.2	1.3	2870	9	US-10-175-743-473	Sequence 473, App
c	109	33.2	1.3	2870	9	US-10-176-488-473	Sequence 473, App
c	110	33.2	1.3	2870	9	US-10-176-492-473	Sequence 473, App
c	111	33.2	1.3	2870	9	US-10-176-747-473	Sequence 473, App
c	112	33.2	1.3	2870	9	US-10-176-750-473	Sequence 473, App
c	113	33.2	1.3	2870	9	US-10-176-985-473	Sequence 473, App
c	114	33.2	1.3	2870	9	US-10-176-987-473	Sequence 473, App
c	115	33.2	1.3	2870	9	US-10-176-991-473	Sequence 473, App
c	116	33.2	1.3	2870	9	US-10-176-992-473	Sequence 473, App
c	117	33.2	1.3	2870	9	US-10-176-993-473	Sequence 473, App
c	118	33.2	1.3	2870	9	US-10-184-638-473	Sequence 473, App
c	119	33.2	1.3	2870	9	US-10-227-884-219	Sequence 219, App
c	120	33.2	1.3	2870	9	US-10-173-695-473	Sequence 473, App

## ALIGNMENTS

```

RESULT 1
US-09-781-693A-1
; Sequence 1, Application US/09781693A
; Publication No. US20030054438A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Tai-Jay
; TITLE OF INVENTION: ANDROGEN RECEPTOR COMPLEX-ASSOCIATED
; FILE REFERENCE: 11709-003001
; CURRENT APPLICATION NUMBER: US/09/781,693A
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 60/262,312
; PRIOR FILING DATE: 2001-01-17
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2580
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-781-693A-1

```

QY	241	TTACAGCAGAAAGGTTTTTGACAACAATTCGTTTCAGGCGACCGACAACATATTTAGTGCA	300
DB	241	TTACAGCAGAAAGGTTTTTGACAACAATTCGTTTCAGGCGACCGACAACATATTTAGTGCA	300
QY	301	AAGTTCTTACCTTTGACAAATGATAACACAGATGTATCTCTGCTCTGGAGATGAGTAATA	360
DB	301	AAGTTCTTACCTTTGACAAATGATAACACAGATGTATCTCTGCTCTGGAGATGAGTAATA	360
QY	361	TTTTATACCAAGCTTTGACAGATGCGAGAAACCAACAGCAANTGCGCAATTTACGTGTCAT	420
DB	361	TTTTATACCAAGCTTTGACAGATGCGAGAAACCAACAGCAANTGCGCAATTTACGTGTCAT	420
QY	421	TATGGAACCTACTTATGACATATGACTGTACCCAAATCACCTTACACTTTCTCTCTTGT	480
DB	421	TATGGAACCTACTTATGACATATGACTGTACCCAAATCACCTTACACTTTCTCTCTTGT	480
QY	481	GGTGAAGATGGAACCTTTAGGTGGTTTGATACCGCATCAAAACTAGCTGCACAAAGAA	540
DB	481	GGTGAAGATGGAACCTTTAGGTGGTTTGATACCGCATCAAAACTAGCTGCACAAAGAA	540
QY	541	GATTGTAAAGATGATATTTAACTGTCGAGCTGCTGCCAGCTGTGTGCTATTTCCG	600
DB	541	GATTGTAAAGATGATATTTAACTGTCGAGCTGCTGCCAGCTGTGTGCTATTTCCG	600
QY	601	CCACCAATACCATATTAACCTTGCCTTGGTTGTTCTGCACGCTCAGTACGAATATATGAT	660
DB	601	CCACCAATACCATATTAACCTTGCCTTGGTTGTTCTGCACGCTCAGTACGAATATATGAT	660
QY	661	CGCGGAATCCTGGGCACAGAGCTACAGGGAATTTATCGAGTGCAGGACTACTCGAATG	720
DB	661	CGCGGAATCCTGGGCACAGAGCTACAGGGAATTTATCGAGTGCAGGACTACTCGAATG	720
QY	721	GTTGCCCGTTTTATTCCCTTCCCATCTTAATAATAAGTCCTGCAGAGTGACATCTCTGT	780
DB	721	GTTGCCCGTTTTATTCCCTTCCCATCTTAATAATAAGTCCTGCAGAGTGACATCTCTGT	780
QY	781	TACAGTGAAGATGGTCAAGAGATTCCTGTTAGTTACTCTTCAGAGTACATATATCTTTT	840
DB	781	TACAGTGAAGATGGTCAAGAGATTCCTGTTAGTTACTCTTCAGAGTACATATATCTTTT	840
QY	841	GACCCCAAGATGATACAGCAGAGAACTTAAACTCCTTCGCGAAGAGAGAAGAA	900
DB	841	GACCCCAAGATGATACAGCAGAGAACTTAAACTCCTTCGCGAAGAGAGAAGAA	900
QY	901	GAGTTCGCAACACCAGTTAAGCGTTTGAGACTTCGTGTTGGTCAGATACCTGGA	960
DB	901	GAGTTCGCAACACCAGTTAAGCGTTTGAGACTTCGTGTTGGTCAGATACCTGGA	960
QY	961	CCCAGAGCAAGCGCGAGAGTCAACGAGAAGATGAGAGCAGAGTCCCAATGTGTCA	1020
DB	961	CCCAGAGCAAGCGCGAGAGTCAACGAGAAGATGAGAGCAGAGTCCCAATGTGTCA	1020
QY	1021	TTGATGCAGAGATGTCGTATATGTTATCAAGATGGTTTGAAGAAGCAAGTGAGTTGCA	1080
DB	1021	TTGATGCAGAGATGTCGTATATGTTATCAAGATGGTTTGAAGAAGCAAGTGAGTTGCA	1080
QY	1081	CAAGCAATAGAGGACGAGGAAGTCTGCACCGAGGTGGAACAAGTCAATCAGATATT	1140
DB	1081	CAAGCAATAGAGGACGAGGAAGTCTGCACCGAGGTGGAACAAGTCAATCAGATATT	1140
QY	1141	TCAACTCTTCCTACGTCGCCATCAAGTCCTGATTTGGAAGTGAGTGAACCTGCAATGGAA	1200
DB	1141	TCAACTCTTCCTACGTCGCCATCAAGTCCTGATTTGGAAGTGAGTGAACCTGCAATGGAA	1200
QY	1201	GTAGATACTCCAGCTCAACAATTTCTCAGCCTTCTACATCCTTACAATGTCAGCTCAG	1260
DB	1201	GTAGATACTCCAGCTCAACAATTTCTCAGCCTTCTACATCCTTACAATGTCAGCTCAG	1260
QY	1261	GCTCATTCGACATCATCTCCACAGAAAGCCCTCATTTACTCTCTTTGCTATCTCTCCA	1320
DB	1261	GCTCATTCGACATCATCTCCACAGAAAGCCCTCATTTACTCTCTTTGCTATCTCTCCA	1320

QY 1321 GACAGTGAACAAAGGCGCTGTTGAGGATCTGTGACACACACACATCATCAGTCTGAT 1380  
DB 1321 GACAGTGAACAAAGGCGCTGTTGAGGATCTGTGACACACACACATCATCAGTCTGAT 1380  
QY 1381 AACATAATGAAAGCTGAGCCCAACAGGACAGGTGAACCACTTTAAGTTTGCAC 1440  
DB 1381 AACATAATGAAAGCTGAGCCCAACAGGACAGGTGAACCACTTTAAGTTTGCAC 1440  
QY 1441 TACAGCAGAGGAGGAACTACAGACACAATAAACTTTACAGATGAATGGAGC 1500  
DB 1441 TACAGCAGAGGAGGAACTACAGACACAATAAACTTTACAGATGAATGGAGC 1500  
QY 1501 AGTATGATCAAGTCTAGAGGATCTGGAGCCATTCGAATCTGAGGGTCAGGAGAA 1560  
DB 1501 AGTATGATCAAGTCTAGAGGATCTGGAGCCATTCGAATCTGAGGGTCAGGAGAA 1560  
QY 1561 TCTTTGCTCCACAGAGCTCAGTCAACACAGGAGAGAGTATCTGCAAGAAAGCTCCT 1620  
DB 1561 TCTTTGCTCCACAGAGCTCAGTCAACACAGGAGAGAGTATCTGCAAGAAAGCTCCT 1620  
QY 1621 GAAGATCATCAGAGGATGTGACAAATATCAGGAAGGAGTATCTGCAAGAAAGCTCCT 1680  
DB 1621 GAAGATCATCAGAGGATGTGACAAATATCAGGAAGGAGTATCTGCAAGAAAGCTCCT 1680  
QY 1681 GAGAACCATATCAATATAACCAATCAGATAAGTTTCACAGCCAGCCATTTGGATTCCAA 1740  
DB 1681 GAGAACCATATCAATATAACCAATCAGATAAGTTTCACAGCCAGCCATTTGGATTCCAA 1740  
QY 1741 TCAGGAGAAAGAAATGACCTCAATCTTGTGATCGCTCTTGTGGGTTCCAGAGAAATCTGCT 1800  
DB 1741 TCAGGAGAAAGAAATGACCTCAATCTTGTGATCGCTCTTGTGGGTTCCAGAGAAATCTGCT 1800  
QY 1801 TCATCTGAAAGCCAGGAAACCAAGAACTTCAGATCAGACTAGCAGTGTGATCTGCTAAC 1860  
DB 1801 TCATCTGAAAGCCAGGAAACCAAGAACTTCAGATCAGACTAGCAGTGTGATCTGCTAAC 1860  
QY 1861 AATGAAATTAACCAATCTGAGCCCTCAGTTCCAAACAGAGCACTGGGCCCTTCAGCT 1920  
DB 1861 AATGAAATTAACCAATCTGAGCCCTCAGTTCCAAACAGAGCACTGGGCCCTTCAGCT 1920  
QY 1921 CATGAAGAAATCCACAGGAGTCTGCTCTTCAGGACACAGATGATGATGATGAC 1980  
DB 1921 CATGAAGAAATCCACAGGAGTCTGCTCTTCAGGACACAGATGATGATGATGAC 1980  
QY 1981 CCAGTCTGATCCAGGTCAGGATATCGAGCAGGACCTGGTGTATGACGCTCTGCTGTT 2040  
DB 1981 CCAGTCTGATCCAGGTCAGGATATCGAGCAGGACCTGGTGTATGACGCTCTGCTGTT 2040  
QY 2041 GCCCGTATTCAGGAGTCTTCAGACGGAGAAAGAAAGGAAAGAAATGGAAGAAATTTGGAT 2100  
DB 2041 GCCCGTATTCAGGAGTCTTCAGACGGAGAAAGAAAGGAAAGAAATGGAAGAAATTTGGAT 2100  
QY 2101 ACTTTGAACATTAAGAGCCGCTAGTAAAAATGGTTTATAAGGCCATCGCACTCCAGG 2160  
DB 2101 ACTTTGAACATTAAGAGCCGCTAGTAAAAATGGTTTATAAGGCCATCGCACTCCAGG 2160  
QY 2161 ACAATGATAAAGAGCCATTTCTGGGTCCTAATCTTGTATGATGATGATGATGATGATGAT 2220  
DB 2161 ACAATGATAAAGAGCCATTTCTGGGTCCTAATCTTGTATGATGATGATGATGATGATGAT 2220  
QY 2221 GGCCACATTTTCATCTGGATCGGCACACTGCTGAGCATTTGATGCTTCTGGAAGCTGAT 2280  
DB 2221 GGCCACATTTTCATCTGGATCGGCACACTGCTGAGCATTTGATGCTTCTGGAAGCTGAT 2280  
QY 2281 AATCATGTGTAACCTGCTGAGCCACACATCTGCTTTGACCCAAATTTAGCCTCATCTGGC 2340  
DB 2281 AATCATGTGTAACCTGCTGAGCCACACATCTGCTTTGACCCAAATTTAGCCTCATCTGGC 2340  
QY 2341 ATAGATTATGACATAAGATCTGCTCACCATTAGAGAGTCAAGGATTTTAAACCGGAA 2400  
DB 2341 ATAGATTATGACATAAGATCTGCTCACCATTAGAGAGTCAAGGATTTTAAACCGGAA 2400  
QY 2401 CTTGCTGATGAAGTTATACTCGAAACGAACACTCATGCTGGAAGAAATAGAAACACCAT 2460

DB 2401 CTTGCTGATGAAGTTATACTCGAAAGCACTCATGCTGGAAGAACTAGAAACCAT 2460  
QY 2461 ACAGTTCCAGCCTCTTCAATGTTGAGGATGTTGGCTTCACTTAATCATATCCGAGCTGAC 2520  
DB 2461 ACAGTTCCAGCCTCTTCAATGTTGAGGATGTTGGCTTCACTTAATCATATCCGAGCTGAC 2520  
QY 2521 CGGTTGGAGGTTGACAGATCAGAGGCTCTGCTCAAGAGAACTGAAATCAGGATGAGAA 2580  
DB 2521 CGGTTGGAGGTTGACAGATCAGAGGCTCTGCTCAAGAGAACTGAAATCAGGATGAGAA 2580

## RESULT 2

US-09-781-693A-3  
; Sequence 3, Application US/09781693A  
; Publication No. US20030054438A1  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Tai-Jay  
; TITLE OF INVENTION: ANDROGEN RECEPTOR COMPLEX-ASSOCIATED  
; FILE REFERENCE: 11709-003001  
; CURRENT APPLICATION NUMBER: US/09/781,693A  
; CURRENT FILING DATE: 2002-07-23  
; PRIOR APPLICATION NUMBER: US 60/262,312  
; PRIOR FILING DATE: 2001-01-17  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 3016  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (18)...(2597)  
US-09-781-693A-3

Query Match 100.0%; Score 2580; DB 9; Length 3016;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2580; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCTCTGGGTCGCTCTTACCCACACCTGTTGGGACGTCGAGAAAGGTCCTTCGGG 60  
DB 18 ATGCTCTGGGTCGCTCTTACCCACACCTGTTGGGACGTCGAGAAAGGTCCTTCGGG 77  
QY 61 CTGAGGACCCGTCCTCCGCTCGGAGTCGCTTACCTGGGAAAGAGAAATTTATCCAAAGA 120  
DB 78 CTGAGGACCCGTCCTCCGCTCGGAGTCGCTTACCTGGGAAAGAGAAATTTATCCAAAGA 137  
QY 121 TTAAGACTTGAAGCAACCCCTTAATGTGATGATGTTGTGTTAATACATCTGTGGAAT 180  
DB 138 TTAAGACTTGAAGCAACCCCTTAATGTGATGATGTTGTGTTAATACATCTGTGGAAT 197  
QY 181 GACACTGGAGAAATATATTTATCTGGCTCAGATGACACCAATTTAGTAATTTAGTAATCT 240  
DB 198 GACACTGGAGAAATATATTTATCTGGCTCAGATGACACCAATTTAGTAATTTAGTAATCT 257  
QY 241 TACAGCAAGAAAGTTTGAACAACATTTCTGAGGACCCAGCAACATATTTAGTGCA 300  
DB 258 TACAGCAAGAAAGTTTGAACAACATTTCTGAGGACCCAGCAACATATTTAGTGCA 317  
QY 301 AGTTCTTACCTGTGCAAAATGATAAAGATGATGATGATGATGATGATGATGATGATGAT 360  
DB 318 AGTTCTTACCTGTGCAAAATGATAAAGATGATGATGATGATGATGATGATGATGATGAT 377  
QY 361 TTTTATACCAAGTTGAGCAAGATGCAAGAACCAACACAGACAAATGCAATTTACGTGAT 420  
DB 378 TTTTATACCAAGTTGAGCAAGATGCAAGAACCAACACAGACAAATGCAATTTACGTGAT 437  
QY 421 TATGGAATCTTATGAGATTTATGACTGTACCCCAATGACCTTACCTTTCTCTCTTGT 480  
DB 438 TATGGAATCTTATGAGATTTATGACTGTACCCCAATGACCTTACCTTTCTCTCTTGT 497  
QY 481 GGTGAAGATGGAATGTTAGTGGTTTGTATACAGGCATCAAACTAGCTGCACAAAGAA 540



; APPLICANT: McCoy, John M.  
 ; APPLICANT: LaValle, Edward R.  
 ; APPLICANT: Collins-Racie, Lisa A.  
 ; APPLICANT: Evans, Cheryl  
 ; APPLICANT: Merberg, David  
 ; APPLICANT: Treacy, Maurice  
 ; APPLICANT: Agostino, Michael J.  
 ; APPLICANT: Steininger II, Robert J.  
 ; APPLICANT: Bowman, Michael R.  
 ; APPLICANT: Spaulding, Vikki  
 ; APPLICANT: Wong, Gordon G.  
 ; APPLICANT: Clark, Hilary  
 ; APPLICANT: Fechtel, Kim  
 ; APPLICANT: Howes, Steven H.  
 ; APPLICANT: Resnick, Richard J.  
 ; APPLICANT: Gulukota, Kamalakar  
 ; APPLICANT: Graham, James R.  
 ; APPLICANT: Genetics Institute, Inc.  
 ; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS  
 ; FILE REFERENCE: GIN 6400  
 ; CURRENT APPLICATION NUMBER: US/09/822,846  
 ; CURRENT FILING DATE: 2001-03-29  
 ; PRIOR APPLICATION NUMBER: 60/195,605  
 ; PRIOR FILING DATE: 2000-04-06  
 ; NUMBER OF SEQ ID NOS: 629  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 200  
 ; LENGTH: 1942  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; US-09-822-846-200

Query Match 59.88; Score 1542.4; DB 9; Length 1942;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 1543; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1037	CTGATATGTTATCAAGATGTTTCAAGNACCAAGTAGAGTTGCGAAGCAATAGAGAGC	1096
DB	1	CTGATATGTTATCAAGATGTTTCAAGNACCAAGTAGAGTTGCGAAGCAATAGAGAGC	60
QY	1097	GAGGAAGATCTCGACCCAGAGGTGGAACAACTCAATCAGATATTTCAACTCTTCCACGG	1156
DB	61	GAGGAAGATCTCGACCCAGAGGTGGAACAACTCAATCAGATATTTCAACTCTTCCACGG	120
QY	1157	TCCCATCAAGTCTGATTTGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT	1216
DB	121	TCCCATCAAGTCTGATTTGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT	180
QY	1217	AACAAATTTCTTCAAGCCTTCTACATCTCTTCAATGTCAGCTCAGCTCAGCTCAGCTCAGCT	1276
DB	181	AACAAATTTCTTCAAGCCTTCTACATCTCTTCAATGTCAGCTCAGCTCAGCTCAGCTCAGCT	240
QY	1277	CTCCACAGAAAGCCCTTACTTACTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT	1336
DB	241	CTCCACAGAAAGCCCTTACTTACTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT	300
QY	1337	AGTCTGTGTAGGATCTGGACACACACATCATCATCATCATCATCATCATCATCATCATCAT	1396
DB	301	AGTCTGTGTAGGATCTGGACACACACATCATCATCATCATCATCATCATCATCATCATCAT	360
QY	1397	TGAGCCCAACACAGGACAGGTGAACACAGTGAACACAGTGAACACAGTGAACACAGTGAACAC	1456
DB	361	TGAGCCCAACACAGGACAGGTGAACACAGTGAACACAGTGAACACAGTGAACACAGTGAACAC	420
QY	1457	CAACTACAGCAACAATAAACTGAACCTGATGATGATGATGATGATGATGATGATGATGATGAT	1516
DB	421	CAACTACAGCAACAATAAACTGAACCTGATGATGATGATGATGATGATGATGATGATGATGAT	480
QY	1517	CTAGAGGAATGGAGCCATTCGAAATCTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTC	1576
DB	481	CTAGAGGAATGGAGCCATTCGAAATCTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTC	540
QY	1577	GCTCAGTGAACACACAGGACAGTGAACACAAAGTCTCTGTAAGAAATCATCAGGAGG	1636

DB	541	GCTCAGTGCAACCCACAGAGGACAGTGAACAAAGAGCTCTCTGAAGATCATCAGAGG	600
QY	1637	ATGTGACAAAATATPCAGGAAGGAGTATCTGCGAAGAAACCCAGTTGAGAACCATATCAATA	1696
DB	601	ATGTGACAAAATATPCAGGAAGGAGTATCTGCGAAGAAACCCAGTTGAGAACCATATCAATA	660
QY	1697	TAACACAATCAGATAAGTTTCACAGCCAGCCATTTGGATTCCAACTCAGGAGAGAAATG	1756
DB	661	TAACACAATCAGATAAGTTTCACAGCCAGCCATTTGGATTCCAACTCAGGAGAGAAATG	720
QY	1757	ACCTCAATCTTGATCGCTCTTGTGGGTTCCAGAAAGATCTGCTTCATCTGAAAAAGCA	1816
DB	721	ACCTCAATCTTGATCGCTCTTGTGGGTTCCAGAAAGATCTGCTTCATCTGAAAAAGCA	780
QY	1817	AGGAACCCAGAACTTCAGATCAGACTAGCAGTGAAGTGTACCAATGAAATTAACACCA	1876
DB	781	AGGAACCCAGAACTTCAGATCAGACTAGCAGTGAAGTGTACCAATGAAATTAACACCA	840
QY	1877	ATCCTGAGCCTCAGTTCCAAACAGAACCCACTGGGCTTCAGCTCATCAGAAACATCCA	1936
DB	841	ATCCTGAGCCTCAGTTCCAAACAGAACCCACTGGGCTTCAGCTCATCAGAAACATCCA	900
QY	1937	CCAGGACTCTGCTCTTCCAGGACACAGATGACAGTGAAGTGTACCAATGAAATTAACACCA	1996
DB	901	CCAGGACTCTGCTCTTCCAGGACACAGATGACAGTGAAGTGTACCAATGAAATTAACACCA	960
QY	1997	GTCCAAAGTATCAGCAGGAGCCTGGTGTATGACGCTCTGCTGCTGCTGCTGCTGCTGCTGCT	2056
DB	961	GTCCAAAGTATCAGCAGGAGCCTGGTGTATGACGCTCTGCTGCTGCTGCTGCTGCTGCTGCT	1020
QY	2057	TCCTCAGACGGAGAAAGAAAGAAATGGAAGATTTGATGATGATGATGATGATGATGATGATGAT	2116
DB	1021	TCATCAGACGGAGAAAGAAAGAAATGGAAGATTTGATGATGATGATGATGATGATGATGATGAT	1080
QY	2117	GGCCGCTAGTAAATGTTTATAAGGCCATCGCAACTCCAGGACCAATGATAAAGAG	2176
DB	1081	GGCCGCTAGTAAATGTTTATAAGGCCATCGCAACTCCAGGACCAATGATAAAGAG	1140
QY	2177	CCAATTTCTGGGCTGCTTAACTTTGATGATGATGATGATGATGATGATGATGATGATGATGAT	2236
DB	1141	CCAATTTCTGGGCTGCTTAACTTTGATGATGATGATGATGATGATGATGATGATGATGATGAT	1200
QY	2237	GGGATCGGCACACTGCTGAGCATTGATGCTTCTGGAAGCTGATAAATCATCTGCTGTAAGT	2296
DB	1201	GGGATCGGCACACTGCTGAGCATTGATGCTTCTGGAAGCTGATAAATCATCTGCTGTAAGT	1260
QY	2297	GCTGAGCCACATCCGTTTGACCCCAATTTAGCCTCATCTGGCATAGATTATGACATAA	2356
DB	1261	GCTGAGCCACATCCGTTTGACCCCAATTTAGCCTCATCTGGCATAGATTATGACATAA	1320
QY	2357	AGATCTGGTCAACATTTAGAGAGTCAAGGATTTTAAACCGAAACTTCTGATGATGATGATGAT	2416
DB	1321	AGATCTGGTCAACATTTAGAGAGTCAAGGATTTTAAACCGAAACTTCTGATGATGATGATGAT	1380
QY	2417	TAACTCGAAACGAACTCATGCTGGAAGAACTAGAAACACCATTTACAGTTCCAGCCTCTT	2476
DB	1381	TAACTCGAAACGAACTCATGCTGGAAGAACTAGAAACACCATTTACAGTTCCAGCCTCTT	1440
QY	2477	TCATGTTGAGGATGTTGGCTTCACTTAATCATATCCGAGCTGACCGTTGGAGGTTGACA	2536
DB	1441	TCATGTTGAGGATGTTGGCTTCACTTAATCATATCCGAGCTGACCGTTGGAGGTTGACA	1500
QY	2537	GATCAGAGGCTCTGCTCAAGGAATGAAATGAGGATGAGGAA	2580
DB	1501	GATCAGAGGCTCTGCTCAAGGAATGAAATGAGGATGAGGAA	1544

RESULT 4  
 US-09-864-761-19286  
 ; Sequence 19286, Application US/09864761  
 ; Patent No. US20020048763A1  
 ; GENERAL INFORMATION:





SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 17295  
LENGTH: 261  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AL031287.3  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.6  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.95  
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.3  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.4  
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 0.81  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.9  
OTHER INFORMATION: EST HUMAN HIT: BE747286.1, EVALUE 0.00e+00  
OTHER INFORMATION: NT HIT: g111425522, EVALUE 0.00e+00  
US-09-864-761-17295

Query Match 10.1%; Score 261; DB 10; Length 261;  
Best Local Similarity 100.0%; Pred. No. 8.1e-67;  
Matches 261; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1118 GTGGAACAAGTCATCAGATATTTCAACTCTTCTCAGGTCCTCAAGTCTGATTTGG 1177  
Db 1 GTGGAACAAGTCATCAGATATTTCAACTCTTCTCAGGTCCTCAAGTCTGATTTGG 60  
QY 1178 AAGTGAGTGAACATGCAATGGAAGTATGATCTCCAGCTGACAAATTTCTTCAGCTTCTA 1237  
Db 61 AAGTGAGTGAACATGCAATGGAAGTATGATCTCCAGCTGACAAATTTCTTCAGCTTCTA 120  
QY 1238 CATCCTTACAAATGTCAGCTCAGGCTCATTCGACATCATCTCCACAGAAAGCCCTCAIT 1297  
Db 121 CATCCTTACAAATGTCAGCTCAGGCTCATTCGACATCATCTCCACAGAAAGCCCTCAIT 180  
QY 1298 CTACTCCTTTGCTATCTCTCCACAGATGACAAAGGAGCTGTGTTGAGGCATCTGGAC 1357  
Db 181 CTACTCCTTTGCTATCTCTCCACAGATGACAAAGGAGCTGTGTTGAGGCATCTGGAC 240  
QY 1358 ACCACACATCATCATCTGCTG 1378  
Db 241 ACCACACATCATCATCTGCTG 261

RESULT 6  
US-09-864-761-2560  
Sequence 2560, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharron G.  
APPLICANT: Hanzel, David R.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aemlica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667

PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SEQ ID NO 2560  
LENGTH: 481  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AL033531.7  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.8  
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.3  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.8  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.99  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.4  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.7  
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.9  
US-09-864-761-2560

Query Match 9.9%; Score 255.8; DB 10; Length 481;  
Best Local Similarity 99.2%; Pred. No. 4.3e-65;  
Matches 257; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1696 ATAACACAATCAGATAAGTTCCACAGCCCAAGCCATTGGATTCCAACTCAGGAGAAAGAAAT 1755  
Db 223 ACAGCACATCAGATAAGTTCCACAGCCCAAGCCATTGGATTCCAACTCAGGAGAAAGAAAT 282  
QY 1756 GACCTCAATCTTGATCGCTCTTGTGGGTTCCAGAGAATCTGCTTCATCTGAAAAAGCC 1815  
Db 283 GACCTCAATCTTGATCGCTCTTGTGGGTTCCAGAGAATCTGCTTCATCTGAAAAAGCC 342  
QY 1816 AAGGAACACAGAACTTCAGATCAGACTAGCAGTACTGAGAGTGTACCAATGAAAAATACACC 1875  
Db 343 AAGGAACACAGAACTTCAGATCAGACTAGCAGTACTGAGAGTGTACCAATGAAAAATACACC 402  
QY 1876 AATCCTGAGCCTCAGTTCCAAACAGAGCCCACTGGCCCTTCAGCTCATGAGAGAACATCC 1935  
Db 403 AATCCTGAGCCTCAGTTCCAAACAGAGCCCACTGGCCCTTCAGCTCATGAGAGAACATCC 462  
QY 1936 ACCAGGACTCTGCTCTTC 1954  
Db 463 ACCAGGACTCTGCTCTTC 481

RESULT 7  
US-09-983-965-58  
Sequence 58, Application US/09983965  
Patent No. US20020137160A1  
GENERAL INFORMATION:  
APPLICANT: Warren, Wesley C.  
APPLICANT: Tao, Nengbing

APPLICANT: Byatt, John C.  
APPLICANT: Mathialagan, Nagappan  
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION  
FILE REFERENCE: 37-21(10297)C  
CURRENT APPLICATION NUMBER: US/09/983,965  
CURRENT FILING DATE: 2001-10-26  
PRIOR APPLICATION NUMBER: US 09/465,231  
PRIOR FILING DATE: 1999-12-15  
PRIOR APPLICATION NUMBER: US 60/113,678  
PRIOR FILING DATE: 1998-12-17  
NUMBER OF SEQ ID NOS: 5912  
SEQ ID NO 58  
LENGTH: 409  
TYPE: DNA  
ORGANISM: Bos taurus  
FEATURE:  
OTHER INFORMATION: Clone ID: 07-BOVMS1-015-Q1-E1-B11  
US-09-983-965-58

Query Match 9.4%; Score 241.8; DB 10; Length 409;  
Best Local Similarity 93.7%; Pred. No. 5.3e-61;  
Matches 252; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 2312 CGTTGACCCAAATTTAGCCCTCTCTGCGATAGATATGACATTAAGATCTGGTCCACAT 2371  
DB 3 CGTTGACCCAAATTTCTCGCTTCATCTGCGATAGATATGACATTAAGATCTGGTCCACAC 62

QY 2372 TAGAAGAGTCAAGGATTTTAAACCGAAACCTGCTGATGAAGTTATTAACCTGAAACGAAAC 2431  
DB 63 TAGAGGATCAAGGATTTTAAACCGAAACCTGCTGATGAAGTTATTAACCTGAAACGAAAC 122

QY 2432 TCATGCTGGAAGAACTAGAAACACCATACAGTTCCAGCCCTCTTTCATGTTGAGGATGT 2491  
DB 123 TCATGCTGGAAGAGACTAGAAATACCATACAGTTCCAGCCCTCTTTCATGTTGAGGATGT 182

QY 2492 TGGCTTCACTTAATCATATCCGAGCTGACCGGTTGGAGGTTGACAGATCAGAAGGCTCTG 2551  
DB 183 TGGCTTCACTGAACCATATTCGAGCTGACCGGTTGGAGGTTGACAGATCAGAAGGCTCTG 242

QY 2552 GTCAAGAGAAATGAAATGAGGATGAGGAA 2580  
DB 243 GTCAGGAGAAATGAAATGAGGATGAGGAA 271

RESULT 8  
US-09-864-761-17296  
Sequence 17296, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Acomica-x-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
SEQ ID NO 17296  
LENGTH: 186  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AL031287.3  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.2  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2  
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.4  
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.4  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.3  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2  
OTHER INFORMATION: EST\_HUMAN HIT: BEI80609.1, EVALUATE 1.00e-100  
OTHER INFORMATION: NT HIT: g111425522, EVALUATE 1.00e-38  
US-09-864-761-17296

Query Match 7.2%; Score 186; DB 10; Length 186;  
Best Local Similarity 100.0%; Pred. No. 9.2e-45;  
Matches 186; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 253 GTTTGTGACAAATTCGTTTCAGGCGCAGCAACATATTTAGTGCAAGTTCTTACT 312  
DB 1 GTTTGTGACAAATTCGTTTCAGGCGCAGCAACATATTTAGTGCAAGTTCTTACT 60

QY 313 TGTACAAATGATAACAGATGTTATCTGCTCTGGAGATGGAGTAATATTTATACCAAC 372  
DB 61 TGTACAAATGATAACAGATGTTATCTGCTCTGGAGATGGAGTAATATTTATACCAAC 120

QY 373 GTTGAGCAAGATGCGAAGCAACCAAGCAAGATGTTATCTGCTCTGGAGATGGAGTAATATTTATACCAAC 432  
DB 121 GTTGAGCAAGATGCGAAGCAACCAAGCAAGATGTTATCTGCTCTGGAGATGGAGTAATATTTATACCAAC 180

QY 433 TATGAG 438  
DB 181 TATGAG 186

RESULT 9  
US-09-864-761-499  
Sequence 499, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Hanzel, David K.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.

APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aeomica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 499  
LENGTH: 467  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AL031287.3  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.6  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.95  
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.3  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.4  
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 0.81  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.9

Query Match 6.2%; Score 160; DB 10; Length 467;  
Best Local Similarity 100.0%; Pred. No. 8.6e-37;  
Matches 160; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1116 AGGTGAACAAAGTCAATCAGATATTTCAACTCTTCTCCAGCTCCCAATCAAGTCTGATT 1175  
|||||  
DB 308 AGGTGAACAAAGTCAATCAGATATTTCAACTCTTCTCCAGCTCCCAATCAAGTCTGATT 367  
|||||

QY 1176 GGAAGTGAGTGAAGTCAATGGAAGTAGATATCTCCAGCTGAAACAATTTCTTCAGCCCTTC 1235  
|||||  
DB 368 GGAAGTGAGTGAAGTCAATGGAAGTAGATATCTCCAGCTGAAACAATTTCTTCAGCCCTTC 427  
|||||

QY 1236 TACATCTCTTACATGTGCTCAGCTCAGGCTCATTCGACATCA 1275  
|||||  
DB 428 TACATCTCTTACATGTGCTCAGCTCAGGCTCATTCGACATCA 467  
|||||

RESULT 10  
US-09-864-761-500  
; Sequence 500, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO  
; FILE REFERENCE: Aeomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 500  
; LENGTH: 468  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AL031287.3  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.2  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2  
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.4  
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.4  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.3  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3

; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2  
US-09-864-761-500

Query Match 5.5%; Score 141; DB 10; Length 468;  
Best Local Similarity 100.0%; Pred. No. 3.5e-31;  
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 251 AGTTTGTGACAAATTCGTTTCAGGGCAGGAGCAACATATTAGTGCAGAAAGTTCTTAC 310  
DB 328 AGTTTGTGACAAATTCGTTTCAGGGCAGGAGCAACATATTAGTGCAGAAAGTTCTTAC 387  
QY 311 CTGTGCAATGATAACACATGTCCTCTCGAGATGAGTAAATATTATACCA 370  
DB 388 CTGTGCAATGATAACACATGTCCTCTCGAGATGAGTAAATATTATACCA 447  
QY 371 ACCTTGAGCAAGATGCAGAAA 391  
DB 448 ACCTTGAGCAAGATGCAGAAA 468

## RESULT 11

US-09-770-791-605  
; Sequence 605, Application US/09770791  
; Patent No. US200200620141

; GENERAL INFORMATION:  
; APPLICANT: Gorlach, Jorn  
; APPLICANT: An, Yong-Qiang  
; APPLICANT: Hamilton, Carol M.  
; APPLICANT: Price, Jennifer L.  
; APPLICANT: Raines, Tracy M.  
; APPLICANT: Yu, Yang  
; APPLICANT: Rameaka, Joshua G.  
; APPLICANT: Page, Amy  
; APPLICANT: Matthew, Abraham V.  
; APPLICANT: Ledford, Brooke L.  
; APPLICANT: Woessner, Jeffrey P.  
; APPLICANT: Haas, William David  
; APPLICANT: Krickler, Maja  
; APPLICANT: Slader, Ted  
; APPLICANT: Davis, Keith R.  
; APPLICANT: Allen, Keith  
; APPLICANT: Hoffman, Neil  
; APPLICANT: Hurban, Patrick

; TITLE OF INVENTION: Expressed Sequences of Arabidopsis

; FILE OF INVENTION: thaliana

; FILE REFERENCE: 2029 (PARA-018PRV)

; CURRENT APPLICATION NUMBER: US/09/770,791

; CURRENT FILING DATE: 2001-01-26

; PRIOR APPLICATION NUMBER: 60/178,480

; PRIOR FILING DATE: 2000-01-27

; NUMBER OF SEQ ID NOS: 999

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 605

; LENGTH: 356

; TYPE: DNA

; ORGANISM: Arabidopsis thaliana

US-09-770-791-605

Query Match 2.8%; Score 71.2; DB 10; Length 356;  
Best Local Similarity 55.7%; Pred. No. 1.2e-10;  
Matches 136; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 109 TTTATCAAGATTAACACTTGAGCAACCCCTTAATGTGATGATGTTGTGTTAATACA 168  
DB 104 TTGTTCCGAGACTTCTCAGGAGCAAGAAATGGAGGACATCAAGCTGTGTTAATGCT 163  
QY 169 ATCTGTGGATGACACTGGAGATATATTTATCTGGCTCAGATGACACCAATTAAGTA 228  
DB 164 TTACGCTGGAACCTCTAATGCTTACCTTTTATATCTGGATCAGATGATTTAAGGATCAAT 223  
QY 229 ATTAGTAATCTTACACGACAAAGGTTTTCACAACAATTCGTCAGGCGACCGAGCAAC 288

DB 224 ATTTGGAAATTAATCCAGTCGGAAGCTTTTGCATTTCTATAGATACAGGCGATAGTCCCAAC 283  
QY 289 ATATTAGTGAAGATTTTACCTTGTAACAAATGATAACAGATTTGATCTGCTCTGGA 348  
DB 284 ATCTTCTGACGAAGTTTGTCCCTGAAACCTCTGACGAGCTTGTGGTATCTGCTGCTGGA 343  
QY 349 GATG 352  
DB 344 GATG 347

## RESULT 12

US-09-983-965-46

; Sequence 46, Application US/09983965

; Patent No. US20020137160A1

; GENERAL INFORMATION:

; APPLICANT: Warren, Wesley C.

; APPLICANT: Tao, Nengbing

; APPLICANT: Byatt, John C.

; APPLICANT: Mathalagan, Nagappan

; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
; FILE OF INVENTION: MUSCLE AND FAT DEPOSITION

; FILE REFERENCE: 37-21(10297)C

; CURRENT APPLICATION NUMBER: US/09/983,965

; CURRENT FILING DATE: 2001-10-26

; PRIOR APPLICATION NUMBER: US 09/465,231

; PRIOR FILING DATE: 1999-12-15

; PRIOR APPLICATION NUMBER: US 60/113,678

; PRIOR FILING DATE: 1998-12-17

; NUMBER OF SEQ ID NOS: 5912

; SEQ ID NO 46

; LENGTH: 257

; TYPE: DNA

; ORGANISM: Bos taurus

; FEATURE:

; OTHER INFORMATION: Clone ID: 06-BOVMS1-017-Q1-E1-B9

US-09-983-965-46

Query Match 2.3%; Score 59; DB 10; Length 257;  
Best Local Similarity 60.2%; Pred. No. 3.9e-07;  
Matches 115; Conservative 0; Mismatches 75; Indels 1; Gaps 1;

QY 2374 GAAGAGTCAGGATTTTAAACCGAAACTTCGTGATGAGTTATACCTCGAAACGAACTC 2433  
DB 65 GAGCTCGAAGAGTTTACCTTGGTCCCTCTCATCGGTCATCARGGGCTTCTACAC 124  
QY 2434 ATGCTGGAAGAACTAGAAACACACATTACAGTTCCAGCCTCTTTTCATGTTGAGGATGTTG 2493  
DB 125 AAGCTGGAAGATACCATTCAGTGGATTTCATTTCCACCCGCTTCACGGGGAAGATCTTG 184  
QY 2494 GTTTCACCTTAATCATATCCGAGCTGACCGGTTGGAGGTTGACAGATCAGAAAGGCTCTGGT 2553  
DB 185 GGGTCACTCGAACATATCCCATTTTGGCGATTGG-CAATTCCAAATTAGGTTGCTCGGT 243  
QY 2554 CAAGAGATGA 2564  
DB 244 CAGAAGACAGA 254

## RESULT 13

US-09-925-300-645

; Sequence 645, Application US/09925300

; Patent No. US20020151681A1

; GENERAL INFORMATION:

; APPLICANT: Craig Rosen,

; APPLICANT: Steve Ruben

; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

; FILE REFERENCE: PA101

; CURRENT APPLICATION NUMBER: US/09/925,300

; CURRENT FILING DATE: 2001-08-10

; PRIOR APPLICATION NUMBER: PCT/US00/05988

; PRIOR FILING DATE: 2000-03-08

; PRIOR APPLICATION NUMBER: 60/124,270

;;  
;; PRIOR FILING DATE: 1999-03-12  
;; NUMBER OF SEQ ID NOS: 1890  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 645  
;; LENGTH: 2679  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: misc feature  
;; LOCATION: (3)  
;; OTHER INFORMATION: n equals a,t,g, or c  
;; NAME/KEY: misc feature  
;; LOCATION: (21)  
;; OTHER INFORMATION: n equals a,t,g, or c  
;; NAME/KEY: misc feature  
;; LOCATION: (24)  
;; OTHER INFORMATION: n equals a,t,g, or c  
;; NAME/KEY: misc feature  
;; LOCATION: (41)  
;; OTHER INFORMATION: n equals a,t,g, or c  
;; NAME/KEY: misc feature  
;; LOCATION: (124)  
;; OTHER INFORMATION: n equals a,t,g, or c  
;; NAME/KEY: misc feature  
;; LOCATION: (128)  
;; OTHER INFORMATION: n equals a,t,g, or c  
;;  
US-09-925-300-645

Query Match 2.1%; Score 54.6; DB 10; Length 2679;  
Best Local Similarity 59.3%; Pred. No. 4.2e-05;  
Matches 112; Conservative 0; Mismatches 74; Indels 3; Gaps 1;  
QY 2193 TAACCTTTGTAAGTAGTGCTGCTGAGCTGATAA---TCATCTGTGTAACCTGCCGACACTGC 2252  
Db 315 TGAGTTTGTGTGAGCGTAGTACTGTGGGCACATCTCTCTGGGAGAAATCATCTCG 374  
QY 2253 TGACCATTTGATGCTCTGGAAGCTGATAA---TCATCTGTGTAACCTGCCGACACTGC 2309  
Db 375 CCAGATTATTCAGTTCATGAGGGGACAGGGAGCGGTGTAACACTGCTTGAGCCCCA 434  
QY 2310 TCCGTTTGACCAATTTAGCCTCATCTGGCATAGATTATGACATAAAGATCTGCTCAC 2369  
Db 435 CCCTCACCTGCTGCTGGCAACCACTGGCGCTAGACCATGATGTGAAGATCTGGGCACC 494  
QY 2370 ATTAGAGA 2378  
Db 495 CACAGCTGA 503

RESULT 14  
US-09-828-447-6  
; Sequence 6, Application US/09828447  
; Patent No. US20020069432A1  
; GENERAL INFORMATION:  
; APPLICANT: COSTA E SILVA, OSWALDO DA  
; APPLICANT: BOHNETT, HANS J.  
; APPLICANT: VAN THIELEN NOCHA  
; APPLICANT: CHEN, ROUYING  
; APPLICANT: ISHITANI, MANABU  
; TITLE OF INVENTION: SIGNAL TRANSDUCTION STRESS-RELATED PROTEINS AND METHODS  
; FILE REFERENCE: 16313-0037  
; CURRENT APPLICATION NUMBER: US/09/828,447  
; PRIOR FILING DATE: 2001-08-20  
; PRIOR APPLICATION NUMBER: 60/196,001  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 2276  
; TYPE: DNA  
; ORGANISM: Physcomitrella patens  
US-09-828-447-6

Query Match 1.6%; Score 41; DB 10; Length 2276;  
Best Local Similarity 47.8%; Pred. No. 0.39;  
Matches 119; Conservative 0; Mismatches 130; Indels 0; Gaps 0;  
QY 1636 GATGTGCAAAATATCAGGAAGAGTATCTGCAGAAAACCCAGTTGAGAACCATATCAAT 1695  
Db 1050 GATAAGTTGCAGGAGCAGTCAACTGCTCTGTTAAAGCCCGCTTGAGAAAAAGATTGCA 1109  
QY 1696 ATAACAATATCAGATAAGTTTCACAGCCAGCCATTGGATTCCCACTCAGGAGAAAGAAAT 1755  
Db 1110 GTTCCACCATCAGAGAAGCAAAATCCATTTCCGAAGAGAGGACTTGAGTGAAAAAGTT 1169  
QY 1756 GACCTCAATCTTGATCGCTCTTGTGGGTTCCAGAAAGAAATCTGCTTCATCTCAAAAAGCC 1815  
Db 1170 GGAATTTTACGTGTTGATTTCAGAGGGTGAATCAGCTGATCTGCCCCCTCAAGTTCCCCC 1229  
QY 1816 AAGGAACCAAACTTCAGATCAGACTAGCAGTGTGCTTACCAATGAAAAATACACC 1875  
Db 1230 CACGGTAAGAACCAATTCAGACGGATAGTGAAGAGTGACGATGACGACAATAAGAG 1289  
QY 1876 AATCCTGAG 1884  
Db 1290 AATCCTGAG 1298

RESULT 15  
US-09-790-988-1  
; Sequence 1, Application US/09790988  
; Patent No. US20020127687A1  
; GENERAL INFORMATION:  
; APPLICANT: SHIGENOBU, SHUJI  
; APPLICANT: WATANABE, HIDEMI  
; APPLICANT: HATTORI, MASAHIRA  
; APPLICANT: SAKAKI, YOSHIYUKI  
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS  
; FILE REFERENCE: 081356/0159  
; CURRENT APPLICATION NUMBER: US/09/790,988  
; PRIOR FILING DATE: 2001-02-23  
; PRIOR APPLICATION NUMBER: JP2000-107160  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 640681  
; TYPE: DNA  
; ORGANISM: Buchnera sp.  
US-09-790-988-1

Query Match 1.5%; Score 39.2; DB 10; Length 640681;  
Best Local Similarity 50.0%; Pred. No. 77;  
Matches 98; Conservative 0; Mismatches 98; Indels 0; Gaps 0;  
QY 1603 AGTCAAAACAAAGCTCCTGGAAGATCATCAGAGGTGACAAAATATCAGAGGAGTA 1662  
Db 198822 AGTCACACTAAACAGAGAAGATATATTCGATACCGCTTAAACAATCATCTTCAGATGAG 198881  
QY 1663 TCTCAGAAAACCCAGCTTGAGAACCATATATACACAATCAGATAAGTTTCACAGCC 1722  
Db 198882 TATACAGAAATATATTTTAAATAATAACACTGATATGATTTTAAATAATCTCATAC 198941  
QY 1723 AAGCCATTGGATTCCAACTCAGGAGAAAGAAATGACCTCAATCTTGATCGCTCTTGGG 1782  
Db 198942 AATGAATCAGATATAAAATTTTAGAGATCAAGTGAAGAAGATTAATAATAATGTTGAT 199001  
QY 1783 GTTCCAGAAGATCTG 1798  
Db 199002 GATACAGACGATATTG 199017

RESULT 16  
US-09-925-300-145  
; Sequence 145, Application US/09925300

Patent No. US20020151681A1  
GENERAL INFORMATION:  
APPLICANT: Craig Rosen,  
APPLICANT: Steve Ruben  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA101  
CURRENT APPLICATION NUMBER: US/09/925,300  
CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05988  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 1890  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 145  
LENGTH: 1945  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1934)  
OTHER INFORMATION: n equals a,t,g, or c  
US-09-925-300-145

Query Match 1.58; Score 38.6; DB 10; Length 1945;  
Best Local Similarity 49.28; Pred. No. 1.8;  
Matches 98; Conservative 1; Mismatches 100; Indels 0; Gaps 0;  
QY 2179 AATTCTGGGGTGTCTAACTTTGTAATGAGTGGTCTGACTGTGCGGCACATTTTCATCTGG 2238  
DB 658 AATTCTATGTCACACAGGCATTTAATCAGTGGAGCGGAGATGCACTCTGTATCTGG 717  
QY 2239 GATCGGCACACATGCTGAGCATTTGATGCTTCTGGAAGCTGATAATCATGTGGTAACTGC 2298  
DB 718 GATCAAGAAATGGGAATGCTCAARTCAATTAAGTCAACAAAGGACAGGTGACCTTC 777  
QY 2299 CTGAGCCACATCGTTGACCCCAATTTAGCCCTCATCTGCGTAGATTTATGACATAAG 2358  
DB 778 CTTTCTATTCACCATCTGGCAAGTTGGCCCTGTGCGTTGGTACAGATAAAACTTTAAGA 837  
QY 2359 ATCTGGTCACCATTTAGAAG 2377  
DB 838 ACGTGAATCTTTGTAGAAG 856

RESULT 17  
US-09-765-272-37  
Sequence 37, Application US/09765272  
Patent No. US20020061545A1  
GENERAL INFORMATION:  
APPLICANT: Choi et. al.  
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
NUMBER OF SEQUENCES: 452  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/765,272  
FILING DATE: 22-Jan-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/961,083  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 37:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1360 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 37:  
US-09-765-272-37

Query Match 1.48; Score 37.4; DB 10; Length 1360;  
Best Local Similarity 44.88; Pred. No. 3.1;  
Matches 143; Conservative 0; Mismatches 176; Indels 0; Gaps 0;  
QY 1611 AAAAGCTCCTGAAGAATCATCAGAGGATGTGACAAAATATCAGGAAGAGTATCTGCAGA 1670  
DB 271 AAAAAGGTAGCAGAGCTGAGAAGAAGTTGAAGAAGCTAAGAAAAAGCCGAGATCA 330  
QY 1671 AAACCCAGTTGAGAACCATATCAATATAACACAAATCAGATAAGTTTCACAGCCAGCCATT 1730  
DB 331 AAAAGAAGAGTCCGCTGTAATCTACCCCAACCAATCTTACAAAACGCTTGAACCTGAAT 390  
QY 1731 GGAATTCACACTCAGGAGAGAATAATGACCTCAATCTTGTGATCGCTCTTGTGGGGTTCCAGA 1790  
DB 391 TGCTGAGTCCGATGTGGAAGTTAAAAAAGCGAGCTTGAACCTAGTAAAAAGAGAACTAA 450  
QY 1791 AGAATCTGCTTCATCTGAAAAGCCAGGACCAAGGAACTTCAGATCAGACTAGCAGCTGA 1850  
DB 451 GGAACCTCGAAACGAGGAAAAAGTTAAGCAAGCAAGCGGAAGTTGAGAGTAAAAAAGC 510  
QY 1851 GAGTGTCTACCAATGAAATAACACCAATCTCTGAGCCTCAGTTCCAAAACAGAGCCACTGG 1910  
DB 511 TGAGCTACAGGTTAGAAAATCAAGACAGATCGTAAAAAAGCAGAGAAGAGCTAA 570  
QY 1911 GCCTTCAGCTCATGAAGAA 1929  
DB 571 ACGAAAGCAGCAGAGAA 589

RESULT 18  
US-09-917-800A-1663/c  
Sequence 1663, Application US/09917800A  
Patent No. US20020119462A1  
GENERAL INFORMATION:  
APPLICANT: Mendrick, Donna  
APPLICANT: Porter, Mark  
APPLICANT: Johnson, Kory  
APPLICANT: Castle, Arthur  
APPLICANT: Elashoff, Michael  
APPLICANT: Gene Logic, Inc.  
TITLE OF INVENTION: Molecular Toxicology Modeling  
FILE REFERENCE: 44921-5038-US  
CURRENT APPLICATION NUMBER: US/09/917,800A  
CURRENT FILING DATE: 2001-07-31  
PRIOR APPLICATION NUMBER: US 60/222,040  
PRIOR FILING DATE: 2000-07-31  
PRIOR APPLICATION NUMBER: US 60/222,880  
PRIOR FILING DATE: 2000-11-02  
PRIOR APPLICATION NUMBER: US 60/290,029  
PRIOR FILING DATE: 2001-05-11  
PRIOR APPLICATION NUMBER: US 60/290,645  
PRIOR FILING DATE: 2001-05-15  
PRIOR APPLICATION NUMBER: US 60/292,336  
PRIOR FILING DATE: 2001-05-22  
PRIOR APPLICATION NUMBER: US 60/295,798  
PRIOR FILING DATE: 2001-06-06  
PRIOR APPLICATION NUMBER: US 60/297,457  
PRIOR FILING DATE: 2001-06-13

	Query Match	1.48;	Score 36.6;	DB 9;	Length 574;
	Best Local Similarity	7.5%;	Pred. NO. 2.9;		
	Matches 36;	Conservative 143;	Mismatches 301;	Indels	0; Gaps
2	TGTCGCGGGTGCCTACCCACACCTGTTGTGGGAGCTGAGGAAAAGTCCCTCGGGC	61			
496	HSB...SN...S.WHB.MTTBKS.B.TS.PS.SBR.CDN.H.AR.AWK..TY....SBM	437			
62	TGGAGGACCCGCTCCGGGCTCGGAGTCCTACCTGGGGAAGAGACAATTTATCCAAAGAT	121			
436	ABSNT.....K...D...YH.T..RD.RS.MN.MCMSA.A.ADBR.Y.RW.BAH...B...	377			
122	TAAACTTGAAGCAACCCCTTAATGTGCATGATGGTTGTGTTAATACAATCTGTGGAATG	181			

Query Match	1.4%	Score 36.6;	DB 9;	Length 671;
Best Local Similarity	8.9%;	Pred. No. 3.2;		
Matches 32;	Conservative 139;	Mismatches 189;	Indels 1;	Gaps 1;
QY	1322	ACATGAACAAAGCGAGTCTGTTGAGGCATCTGGACACACACACATCATCAGCTCTGATA	1381	
Db	51	AFLGPKDLFPYDKCKYKPKNRKGFNEGLWETQNNPHASYSA PPPVSSSDSEAPRNP	110	
QY	1382	ACAAATGAACAGCTGAGCCCCAAACAGGACGAGTGTAACCGAGTTTAAAGTTTGCACT	1441	
Db	111	AGSDADEDEDEDGVMVAVTAVTATAASDRMESDSDSKSSDNGLKRTTALAKMSVSKRA	170	
QY	1442	ACAGCACAGAAGGA-ACAAC TACAAGCAACAATAAACTGAAC TTTACAGATGAATGAGGC	1500	
Db	171	RKASSDLQASVSPSEENSESSEKTSQDQFTPEKKAIVAPRPGPLGGKKKKAPS	230	
QY	1501	AGTATAGCATCAAGTCTAGAGGAATTTGGAGCCATTCGAAATCTGAGGGTTCAGGAGGAA	1560	
Db	231	ASDSDSKAOSDCAKPPVAMARAKSSSSSSSSSDSDSVKPPGPKPAEKLPPKPRG	290	

Downloaded from <http://ajphaphysocpharm.sagepub.com/> at 11:00 11 September 2014

\_\_\_\_\_





;; PRIOR FILING DATE: 1999-05-14  
;; PRIOR APPLICATION NUMBER: PCT/NZ99/00051  
;; PRIOR FILING DATE: 1999-04-29  
;; PRIOR APPLICATION NUMBER: 09/188,930  
;; PRIOR FILING DATE: 1998-11-09  
;; PRIOR APPLICATION NUMBER: 09/069,726  
;; PRIOR FILING DATE: 1998-04-29  
;; NUMBER OF SEQ ID NOS: 725  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 618  
;; LENGTH: 528  
;; TYPE: DNA  
;; ORGANISM: Human  
US-10-152-661-618

Query Match 1.4%; Score 35.4; DB 9; Length 528;  
Best Local Similarity 50.9%; Pred. No. 6.1;  
Matches 84; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 1118 GTGGAACAAGTCAATCAGATATTTCCTTCTAGGGTCCCATCAAGTCTGATTGG 1177  
DB 64 GGAATCAAGACATAGTGGCTTCTGCTAAGCCAGGGCTGCCACAATGACACAGTAGC 123

QY 1178 AAGTGAAGTGAAGTCAATGGAAGTAGATCTCCAGCTGAACAATTTCTTCAGCCTTCTA 1237  
DB 124 CAGATCTGCAATTCCTCAATGAGAGCAGCCCAATACAGAAGAAAGCATCGCACATCGTTGC 183

QY 1238 CATCTCTACAATGTCTAGCTCAGGCTCATTCGACATCATCTCCCA 1282  
DB 184 CACAGTTTTCATGGCCACAGCCACAGCAGCATCTCTGTCCA 228

RESULT 26  
US-09-866-050A-618  
; Sequence 618, Application US/09866050A  
; Publication No. US2003004071A1

;; GENERAL INFORMATION:  
;; APPLICANT: Watson, James D.  
;; APPLICANT: Strachan, Lorna  
;; APPLICANT: Sleeman, Matthew  
;; APPLICANT: Onrust, Rene  
;; APPLICANT: Murison, James G.  
;; APPLICANT: Kumble, Krishanand D.  
;; TITLE OF INVENTION: Compositions Isolated From Skin Cells  
;; FILE OF INVENTION: and Methods for Their Use  
;; FILE REFERENCE: 11000.101ic4U  
;; CURRENT APPLICATION NUMBER: US/09/866,050A  
;; CURRENT FILING DATE: 2001-05-24  
;; NUMBER OF SEQ ID NOS: 725  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 618  
;; LENGTH: 528  
;; TYPE: DNA  
;; ORGANISM: Human  
US-09-866-050A-618

Query Match 1.4%; Score 35.4; DB 9; Length 528;  
Best Local Similarity 50.9%; Pred. No. 6.1;  
Matches 84; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 1118 GTGGAACAAGTCAATCAGATATTTCCTTCTAGGGTCCCATCAAGTCTGATTGG 1177  
DB 64 GGAATCAAGACATAGTGGCTTCTGCTAAGCCAGGGCTGCCACAATGACACAGTAGC 123

QY 1178 AAGTGAAGTGAAGTCAATGGAAGTAGATCTCCAGCTGAACAATTTCTTCAGCCTTCTA 1237  
DB 124 CAGATCTGCAATTCCTCAATGAGAGCAGCCCAATACAGAAGAAAGCATCGCACATCGTTGC 183

QY 1238 CATCTCTACAATGTCTAGCTCAGGCTCATTCGACATCATCTCCCA 1282  
DB 184 CACAGTTTTCATGGCCACAGCCACAGCAGCATCTCTGTCCA 228

## RESULT 27

US-09-925-301-690/c  
; Sequence 690, Application US/09925301  
; Patent No. US20020052308A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA106  
; CURRENT APPLICATION NUMBER: PCT/US00/05882  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: 2000-03-08  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 1694  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 690  
; LENGTH: 594  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (2)  
; OTHER INFORMATION: n equals a.t.g, or c  
; NAME/KEY: misc\_feature  
; LOCATION: (473)  
; OTHER INFORMATION: n equals a.t.g, or c  
US-09-925-301-690

Query Match 1.4%; Score 35.4; DB 10; Length 594;  
Best Local Similarity 50.9%; Pred. No. 6.7;  
Matches 84; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 1118 GTGGAACAAGTCAATCAGATATTTCCTTCTAGGGTCCCATCAAGTCTGATTGG 1177  
DB 468 GGAATCAAGACATAGTGGCTTCTGCTAAGCCAGGGCTGCCACAATGACACAGTAGC 409

QY 1178 AAGTGAAGTGAAGTCAATGGAAGTAGATCTCCAGCTGAACAATTTCTTCAGCCTTCTA 1237  
DB 408 CAGATCTGCAATTCCTCAATGAGAGCAGCCCAATACAGAAGAAAGCATCGCACATCGTTGC 349

QY 1238 CATCTCTACAATGTCTAGCTCAGGCTCATTCGACATCATCTCCCA 1282  
DB 348 CACAGTTTTCATGGCCACAGCCACAGCAGCATCTCTGTCCA 304

## RESULT 28

US-09-770-445-480  
; Sequence 480, Application US/09770445  
; Patent No. US20020023281A1

;; GENERAL INFORMATION:  
;; APPLICANT: Gorlach, Jorn  
;; APPLICANT: An, Yong-Qiang  
;; APPLICANT: Hamilton, Carol M.  
;; APPLICANT: Price, Jennifer L.  
;; APPLICANT: Raines, Tracy M.  
;; APPLICANT: Yu, Yang  
;; APPLICANT: Rameaka, Joshua G.  
;; APPLICANT: Page, Amy  
;; APPLICANT: Matthew, Abraham V.  
;; APPLICANT: Ledford, Brooke L.  
;; APPLICANT: Woessner, Jeffrey P.  
;; APPLICANT: Haas, William David  
;; APPLICANT: Garcia, Carlos A.  
;; APPLICANT: Krickler, Maja  
;; APPLICANT: Slader, Ted  
;; APPLICANT: Davis, Keith R.  
;; APPLICANT: Allen, Keith  
;; APPLICANT: Hoffman, Neil  
;; APPLICANT: Hurlan, Patrick  
;; TITLE OF INVENTION: Expressed Sequences of Arabidopsis  
;; TITLE OF INVENTION: thaliana  
;; FILE REFERENCE: 2023US (PARA-012PRV)

	Query Match	1.4%	Score 35.4	DB 10	Length 1548
	Best Local Similarity	50.9%	Pred. No. 13		
	Matches 84	Conservative 0	Mismatches 81	Indels 0	Gaps
QY 1118	GTGGAAACAAGTCAATCAGATATTTCAACTCTTCCTACGGTCCCATCAACTCTGATTGG	1177			
Db	455 GGGNAATCAACACATAGTGGTCTTCTGTGACCAAGGGCTGCCACAATGACACAGTAGC	396			
QY 1178	AAGTGAAGTGAACATGCNAATGGAAGTAGATACTCCAGCTGAACAATTTCTTCACGCCTTCTA	1237			

Db 395 CAGATCTGCAATTCGAATGAGAGCAGCCCAATACAGAGAAAGCATCGCACATGCTTTC 336  
Qy 1238 CATCTCTCAAAATGCTAGCTCAGCTCATTCGACATCATCTCCCA 1282  
Db 335 CACAGTTTTCATGGCCACAGACGCCACAGCAGTCTCTGTTCCA 291

## RESULT 32

US-09-801-368-33  
; Sequence 33, Application US/09801368  
; Patent No. US20020128250A1  
; GENERAL INFORMATION:  
; APPLICANT: Busby, Robert  
; APPLICANT: Cali, Brian  
; APPLICANT: Hecht, Peter  
; APPLICANT: Holtzman, Doug  
; APPLICANT: Madden, Kevin  
; APPLICANT: Maxon, Mary  
; APPLICANT: Milne, Todd  
; APPLICANT: No. US20020128250A1man, Thea  
; APPLICANT: Royer, John  
; APPLICANT: Salama, Sofie  
; APPLICANT: Sherman, Amir  
; APPLICANT: Silva, Jeff  
; APPLICANT: Summers, Eric  
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi  
; FILE REFERENCE: 109272.147  
; CURRENT FILING DATE: 2001-03-07  
; PRIOR APPLICATION NUMBER: US 09/487,558  
; PRIOR FILING DATE: 2000-01-19  
; PRIOR APPLICATION NUMBER: US 60/160,587  
; PRIOR FILING DATE: 1999-10-20  
; NUMBER OF SEQ ID NOS: 440  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 33  
; LENGTH: 3972  
; TYPE: DNA  
; ORGANISM: Saccharomyces cerevisiae  
US-09-801-368-33

Query Match 1.4%; Score 35.2; DB 10; Length 3972;  
Best Local Similarity 46.7%; Pred. No. 30;  
Matches 112; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

Qy 408 ATTACGCTGCTATTGGAAGTACTTATGAGATTATGACTGTACCAATGACCCCTTACAC 467  
Db 3417 ATTTAGGCAACAATTTATAAATAAGATAATGGTATGCTTGTCCCTGCTTTATCGCAA 3476  
Qy 468 TTTTCTCTCTTGTGGTGAAGATGGAAGTGTAGGTGGTTTATACACACGCATCAAACTAG 527  
Db 3477 TGAACACCATATGTTGCAAGTGCAGCTGTTAAGACTAAGTTAGCCAAAAGATCAATGT 3536  
Qy 528 CTGCACAAAGAAGATTGTAAGATGATATTTTAATTAACGTGCGACGTGCTGCCACGTC 587  
Db 3537 TGAAGGGCTTCAATGTTTATTAATGAATCCTAGTAACTCTTTTACGATACCTCTTT 3596  
Qy 588 TGTGCTATTTGCCACCAATACCATATTAACCTGCTGTTGTTGTTCTGACAGCTCAGT 647  
Db 3597 TTTGAATATGAGGATCCTATTTCGAATGAATTTTCCCTTTGATATGGGACAGGGCAGT 3656

## RESULT 33

US-09-860-432-73  
; Sequence 73, Application US/09860432  
; Patent No. US20020081606A1  
; GENERAL INFORMATION:  
; APPLICANT: TRIEU-CUOT, Patrick  
; APPLICANT: POHART, Claire  
; TITLE OF INVENTION: METHODS FOR DETECTING AND IDENTIFYING GRAM POSITIVE BACTERIA IN A  
; FILE REFERENCE: 206427USO  
; CURRENT APPLICATION NUMBER: US/09/860,432  
; CURRENT FILING DATE: 2001-09-12

; PRIOR APPLICATION NUMBER: 60/205,237  
; PRIOR FILING DATE: 2000-05-19  
; NUMBER OF SEQ ID NOS: 96  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 73  
; LENGTH: 429  
; TYPE: DNA  
; ORGANISM: Staphylococcus hyicus  
US-09-860-432-73

Query Match 1.4%; Score 35; DB 10; Length 429;  
Best Local Similarity 49.7%; Pred. No. 6.9;  
Matches 89; Conservative 0; Mismatches 90; Indels 0; Gaps 0;  
Qy 1043 TGTATCAAGATGTTTGAAGAAGCAAGTGGTGTGCACAAAGCAATAGAGAGCAGGAA 1102  
Db 209 TTTTATCCCCAGAAATCTGAAGAAAAAGTGAAGTTGTTGACAAAAATTAAGAACAAATGGG 268  
Qy 1103 GATCTGACCCAGAGGTGGGAACAAGTCAATCAGATATTTCAACTCTTCTTACGGTCCCAT 1162  
Db 269 GTTCTTTAGACGCAATTTAAAAAAGAAATTTCTCAGATAAAGCAGCAGCAGGATTTGGATCTG 328  
Qy 1163 CAAGTCTCTGATTTGGAAAGTGAAGTCAACTGCAATGGAAGTAGACTCCAGCTGAACAA 1221  
Db 329 GCTGGGCTTGCTGTAGTAATAATGCTCAATTAGAAATTTGTACACACACCAACCA 387

## RESULT 34

US-09-998-598-61/c  
; Sequence 61, Application US/09998598  
; Patent No. US20020150922A1  
; GENERAL INFORMATION:  
; APPLICANT: Stolk, John A.  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Chenault, Ruth A.  
; APPLICANT: Meagher, Madelein Joy  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER  
; FILE REFERENCE: 210121.561  
; CURRENT APPLICATION NUMBER: US/09/998,598  
; CURRENT FILING DATE: 2001-11-16  
; NUMBER OF SEQ ID NOS: 2606  
; SOFTWARE: Corixa Invention Disclosure Database  
; SEQ ID NO 61  
; LENGTH: 453  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-998-598-61

Query Match 1.4%; Score 35; DB 10; Length 453;  
Best Local Similarity 50.9%; Pred. No. 7.2;  
Matches 83; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

Qy 1120 GGAACAAGTCAATCAGATATTTTCAACTCTTCTCAGGTCCCATCAAGTCCCTGATTGGAA 1179  
Db 453 GAATCAAGACATAGTGGTCTTCTCTAAGCAAGGCTGCCACAATGACACAGTAGCCA 394  
Qy 1180 GTGAGTGAATGCAATGGAAGTAGATTAATCTCAGCTGAACAATTTCTTCAGCCTTCTACA 1239  
Db 393 GATCTGCAATTTCCAAATGAGAGCAGCCCAATACAGAAAGAACATCGCACATGTTTGCCA 334  
Qy 1240 TCCTCTACAATGTCAGCTCAGGCTCATTCGACATCATCTCCCA 1282  
Db 333 CAGTTTTCATGGCCACAGCAGCCACAGCATCTCTGTTCCA 291

## RESULT 35

US-09-864-761-7684/c  
; Sequence 7684, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.



Query Match 1.4%; Score 35; DB 9; Length 6823;  
Best Local Similarity 48.7%; Pred. No. 52;  
Matches 95; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

QY 1287 AAGCCCTCACTCTACTCTTCTGCTATCTTCTCCAGACAGTGAACAAAGGCGAGTCTGTGA 1346  
DB 5321 AAACCCCTCAACCTAAACCTGGGTTGGGCTCTAAAAGTAATTTAGGGAGAGCATCA 5262  
QY 1347 GGCATCTGGACACACACATCATCTAGTCTGATAACAATAATGAAAGCTGAGCCCAA 1406  
DB 5261 AGACCTGATTATCAGTCTCACATGGAATCTCAACAGATTTCAAAGCAACTCACCAC 5202  
QY 1407 ACCAGGACAGGTGAACAGTCTTAAAGTTTGCATCTACAGCACAGGAACAACTACAG 1466  
DB 5201 AAGTGTGAGACACAAAAGATCTTAAATAATCATCTAGCAAACTACACTAAGAAGCTCAATG 5142  
QY 1467 CACAATAAACTGAA 1481  
DB 5141 ACCAAAATAATCAA 5127

RESULT 38  
US-09-867-550-1267/c  
; Sequence 1267, Application US/09867550  
; Patent No. US20020082206A1  
; GENERAL INFORMATION:  
; APPLICANT: Leach, Martin D.  
; APPLICANT: Mehraban, Fuad,  
; APPLICANT: Conley, Pamela  
; APPLICANT: Law, Debbie  
; APPLICANT: Topper, James  
; TITLE OF INVENTION: No. US20020082206A1 Polynucleotides from Atherogenic Cells and  
; FILE REFERENCE: 21402-013 (Cura-313)  
; CURRENT APPLICATION NUMBER: US/09/867,550  
; CURRENT FILING DATE: 2001-09-20  
; PRIOR APPLICATION NUMBER: USSN 60/208,427  
; PRIOR FILING DATE: 2000-05-30  
; NUMBER OF SEQ ID NOS: 2125  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1267  
; LENGTH: 929  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)  
; OTHER INFORMATION: Wherein n is one of a or t or c or g  
US-09-867-550-1267

Query Match 1.3%; Score 34.8; DB 10; Length 929;  
Best Local Similarity 45.8%; Pred. No. 14;  
Matches 120; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

QY 1747 GAAAGAAATGACCTCAATCTTGATGCTCTTGTGGGTTCCAGAGAAATCTGCTTCATCT 1806  
DB 501 GAAAGCCGAGGCTGACCAAGACCCAGGTGGGATCATATAAGAGAGCGGGTGGCAGCC 442  
QY 1807 GAAAGAGCCAGGACCAAGAACTTCAGATCAGATAGCTAGAGTGCTACCAATCAA 1866  
DB 441 CAGGATGTTAGACCGCGGAGATGTTGATCCGTTGAATCTCGCAAGACCTTTTGGAGAA 382  
QY 1867 AATACACAAATCTTGAGCCCTCAATCTTCCAAACAGAACCACTGGGCCCTTCAGCTCATGAA 1926  
DB 381 AATGTTCTGAAGCAGAGACAGCAGCGGTGGCGGCGGCTGAGAGTCCCAATCATGTC 322  
QY 1927 GAACATCCACAGGAGTCTGCTCTTCCAGACACAGATGACAGTGTGATGATGACCCAGTC 1986  
DB 321 AAAAGACAACCTCGGTGACGGGTGGCGAGCAGGACACCGCTGATGATGGGATGAGTGACAA 262  
QY 1987 CTGATCCAGGTGCAAGGTATC 2008

Db 261 GTATACCTTGGTGTCTGCTTC 240  
RESULT 39  
US-09-864-761-19041/c  
; Sequence 19041, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO  
; FILE REFERENCE: Aesmica-x-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 19041  
; LENGTH: 3953  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AP00511.1  
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.98  
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.4  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.98  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1  
; OTHER INFORMATION: EST\_HUMAN HIT: AW867076.1, EVALUE 0.00e+00  
; OTHER INFORMATION: SWISSPROT HIT: Q02817, EVALUE 5.00e-63  
; OTHER INFORMATION: NT HIT: AE001609.1, EVALUE 6.50e-01  
US-09-864-761-19041

```

Query Match      1.3%  Score 34.8; DB 10; Length 3953;
Best Local Similarity 51.3%; Pred. No. 40;
Matches 81; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 1786 CCAGAAGAATCTGCTTCATCTGAAAAAGCCAAAGCAACCCAGAACTTCAGATCAGACTAGC 1845
      ||||| || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3309 CCAGAAGAGTCAGAAAAAACTGAAGATTCAGAACACACAGTTGCTTCAGACAGCTCTCTG 3250
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 1846 ACTGAGAGTGCTACCAATGAAAAATPAACCAATCTTGAGCTCAGTTTCCAAACAGAGGCC 1905
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3249 ACAAATAACTAAAAACATACAAGAGACCATATCAGCCAATGAGCTCACACAATCTCTA 3190
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 1906 ACTGGGGCTTCAGCTCATGAGAAACATCCACCAGGGA 1943
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3189 GCAGAGCCTACAGAACATGGAGGAAGGACAGCCAATGA 3152
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 40
US-09-070-927A-210/C
; Sequence 210, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
;              Patrick J. Dillon
;              Steven Barash
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,927A
; FILING DATE: 04-May-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB369
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 210:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7287 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 210:
;
; US-09-070-927A-210

```

Search completed: March 26, 2003, 05:37:07  
Job time : 916 secs

